Universidade de Lisboa
Faculdade de Ciências
Departamento de Biologia Vegetal

# Evolution of *Escherichia coli* in the mouse gut

Marta Mansos Lourenço

Dissertação

Mestrado em Microbiologia Aplicada

2013/2014

Universidade de Lisboa
Faculdade de Ciências
Departamento de Biologia Vegetal

# Evolution of *Escherichia coli* in the mouse gut

Dissertação orientada por
Doutora Ana Margarida Sousa
Doutor Francisco Dionísio

Marta Mansos Lourenço

Mestrado em Microbiologia Aplicada

2013/2014

FUNDAÇÃO CALOUSTE GULBENKIAN

Instituto Gulbenkian de Ciência

# Evolution of *Escherichia coli* in the mouse gut

Marta Mansos Lourenço

Master thesis

2013/2014

This thesis was fully performed at Instituto Gulbenkian de Ciência under the direct supervision of Drª Ana Margarida Sousa and Prof. Dr. Francisco Dionísio was the internal designated supervisor in the scope of the Master in Applied Microbiology of the Faculty of Sciences of the University of Lisbon.

## Acknowledgments

I would like to express my gratitude to all the people who helped me through the learning process of this master thesis. I am deeply grateful to my supervisor, Drª Ana Margarida Sousa, for the useful comments, remarks, data analysis help and knowledge preparing me for future challenges. Thank you for the patience, motivation and persistence during this year. I would also like to express my gratitude to Drª Isabel Gordo for the hypothesis of working in her lab in one of the best Science Institutes in Portugal. Thank you for introducing me the topic and for the helpful discussions and expertise.

I thank all of my lab mates in the Evolutionary Biology Lab for the helpful discussions and joy making the lab a really great place to work. In particular I am also grateful to Catarina Bourgard and Migla Miskinyte for helping me in the beginning of this journey, Catarina Pinto and Daniela Zwerschke for the motivation and support in the lab work, João Batista for all the help and knowledge during the experiment, and Jorge for the patience and assistance in the computational work.

I would like to thank my family and my boyfriend Gonçalo for supporting me throughout entire process by keeping me balanced and motivated and by always supporting my choices.

The work presented in this thesis was funded by: project PFE-GI-UE-ERC-2010-StG-260421 and project PTDC/BIA-EVF/118075/2010.

# Table of Contents

## Summary

The role of bacterial diversity is not completely understood. Several factors can shape this diversity, either within a host or in an external environment. It is important to understand this process and its driving forces both in ecological and evolutionary terms. In this thesis, we study bacterial adaptation to the mammalian gut, which is one of the most complex bacterial environments. The mammalian gut is also thought to be variable over time. To thrive in an environment such as this, the accumulation of adaptive mutations is essential. In populations where several beneficial mutations segregate simultaneously, many small effect mutations are lost due to competition with the ones of larger effect, a process known as clonal interference (CI).

As an adaptive walk is expected to involve more than one adaptive step, we have studied the adaptive mutation corresponding to the first and second steps of adaptation of *Escherichia coli* to the mouse gut. We observed an intense process of clonal interference occurring during the first colonization that seems to decrease in the second colonization. This analysis showed that the rate of adaptation of *E.coli* to the mouse gut seems to be declining as the population adapts. Despite the smaller effect of interference in the second colonization, the comparison of the genetic basis of adaptation from both colonizations revealed a remarkable parallelism (mainly IS driven) in the adaptation of *E.coli* to the mouse gut which continues to adapt in a second colonization to the gut.

We further tested the effect of these mutations in the expression of their respective genes and in the mutant's fitness when exposed to different carbon sources. The results indicate that oxygen level and the metabolic pressures are important stimulus in the adaptation to the studied environment.

In sum, though the gut is a highly complex environment, the repeatability of evolution shows that parallel evolution is not restricted to the laboratory environments and may be very common in nature.

**Contributions to this work:** I performed all the work presented in this thesis, except the work presented in Figure 4, Figure 5 (results from first colonization), Figure 6 (sequencing analysis of the clones) and Figure S1.

Some of the results here presented are already accessible in the *PLoS Genetics* paper "João Barroso-Batista, Ana Sousa, **Marta Lourenço**, Marie-Louise Bergman, Daniel Sobral, Jocelyne Demengeot, Karina B. Xavier, Isabel Gordo. The first steps of adaptation of *Escherichia coli* to the gut are dominated by soft sweeps. PLoS Genet. 2014 Mar;10(3):e1004182.".

This paper was mentioned by Carl Zimmer on an article about "evolutionary forecasting" (http://www.simonsfoundation.org/quanta/20140717-the-new-science-of-evolutionary-forecasting/).

**Keywords:** *Escherichia coli*; Parallel adaptation; Natural environment; Microbiota; Generalists; Molecular Evolution; Insertion Sequences (IS); Gene expression studies

## Sumário

Existe ainda algum desconhecimento sobre o papel da diversidade bacteriana. São muitos os factores que a podem modular, quer num hospedeiro quer num ambiente externo. É importante perceber este processo e as suas forças motrizes tanto em termos evolutivos como ecológicos.

Nesta tese abordamos o estudo da adaptação bacteriana ao intestino dos mamiferos, um dos ambientes bacterianos mais complexos. Julga-se que o ambiente dentro do intestino seja muito variável devido a vários factores como os niveis de oxigénio (flutuantes), os nutrientes ingeridos pelo hospedeiro, o pH, o uso de antibióticos e o sistema imunitário. Para prosperar em ambientes como este, a acumulação de mutações benéficas é essencial. Em populações em que diversas mutações benéficas surgem em simultâneo, as de menor perder-se-ão devido á competição com as de maior efeito que se irão fixar. Se esta mutações forem de impacto semelhante será dificil a fixação de uma delas devido a intensa competição entre elas. A este processo dá-se o nome de interferência clonal.

O organismo modelo que usámos durante este estudo foi a *Escherichia coli* MG1655. É uma bacteria anaeróbia facultativa, gram-negativa e com crescimento óptimo à temperatura de 37ºC. É um dos organismos modelos mais bem estudados que para além do crescimento rápido em condições laboratoriais, tem uma base genética bem estabelecida e uma sequência genómica determinada.

Embora a *E.coli* represente menos de 1% das bactérias presentes no intestino humano (1), está descrita como sendo uma das primeiras a colonizar recém-nascidos (primeiras 40 horas após o parto) e ainda recentemente foi descrita como sendo a bactéria mais abundante na placenta humana (2)(3).

Alem do mais a *E.coli* sendo uma bacteria comensal tem a capacidade de se tornar patogénica. Assim torna-se a candidata ideal para estudos de transição nas interacções bacteria-hospedeiro podendo variar entre mutualismo, comensalismo ou até patogenese oportunista (4).

Nesta tese procedemos à análise da base genética da adaptação da *E.coli* MG1655 durante duas colonizações consecutivas do intestino do ratinho. Para isso foi introduzida por “gavage” (que consiste na introdução directa de bactérias no estômago do ratinho por via de um tubo) uma co-cultura de duas estirpes geneticamente iguais com a excepção da fluorescência codificada no seu genoma (5). Como ancestral da segunda colonização foi usado um dos clones isolados das populações evoluidas durante a primeira colonização. A base genética dos clones evoluídos nestas duas colonizações foi descrita com base em dados de sequenciação do genoma completo (WGS – whole genome sequencing).

A análise genética dos clones mostrou um grande paralelismo entre as mutações adaptativas que surgiram na primeira e segunda colonizações. O primeiro passo adaptativo correspondeu à inactivação do operão *gat* ( responsável pelo metabolismo do galactitol). O segundo passo caracterizou-se pela a inserção de uma sequência de inserção (IS) nas regiões regulatórias de quatro genes o *dcuB*, *focA*, *arcA* e o *yjjP*, pela inserção de um IS na região codificante dos genes *radA* e *oppB* e ainda por uma mutação pontual (ou pequenas deleções ou inserções) no gene *srlR*. Sendo alguns destes genes de elevada importância para a bactéria (por exemplo o gene *arcA* é um dos reguladores mais importantes na passagem de respiração aeróbia para anaeróbia) questionou-se qual seria o papel dos elementos IS na região regulatória destes genes. Estes IS poderiam inactivar o gene, aumentar ou diminuir a sua expressão. Para investigar esta questão medimos o efeito da presença dos elementos IS na expressão dos respectivos genes. Algumas demonstraram possuir um efeito significativo na expressão dos genes, sendo este dependente da presença de oxigénio no meio de cultura, apoiando a possibilidade de os niveis de oxigénio serem uma das pressões selectivas mais importantes dentro do intestino. O efeito destas mutações revelou ainda que os elementos IS têm um papel fundamental neste processo adaptativo a que a *E.coli* está a ser sujeita.

Para além dos genes alvo de inserções medimos ainda a expressão do gene *srlA,* uma vez que este gene faz parte do operão responsável pelo metabolismo do sorbitol estando sob controlo directo do repressor *srlR* (um dos genes mais atingido durante as duas colonizações). Verificámos que o gene *srlA* teve um aumento de expressão tanto em aerobiose como em anaerobiose demostrando assim que o sorbitol é provavelmente uma fonte de carbono importante para a *E.coli* no intestino.

Para melhor perceber as dinâmicas desta adaptação da *E.coli* ao intestino do ratinho determinámos ainda a diversidade de haplótipos ao longo do período adaptativo de algumas das populações das duas colonizações. A análise desta diversidade revelou um regime de interferência clonal em ambas as colonizações (mostrando um dinâmica de “soft sweeps” que consiste no aparecimento de várias mutações ao mesmo tempo competindo entre elas mas nunca conseguindo a fixação na população). Esta interferência aparenta ser inferior na segunda colonização pois o número de haplótipos com mais de uma mutação a surgirem durante a primeira colonização é muito superior aos número dos da segunda colonização. Este estudo revelou ainda a presença de uma população com uma dinâmica adaptativa diferente em que uma das mutações do segundo passo de adaptação se fixou (evidenciando uma dinâmica de “hard sweep” que consiste no aparecimento de uma mutação benéfica que se fixa na população). Estas observações levam à conclusão de que a adaptação parece estar a diminuir apresentando ainda dinâmicas de adaptação diferentes das representadas na primeira colonização.

Para além disso estudámos ainda as potenciais pressões metabólicas presenciadas pela *E.coli* no interior do intestino. Para isso realizámos competições *in vitro* na presença de um conjunto de diferentes fontes de carbono a baixas e altas concentrações no meio de cultura. As fontes de carbono usadas foram o gluconato, o glucuronato, a manose e a ribose descritos como importantes para a *E.coli* quando dentro do intestino. O sorbitol foi tambem utilizado pois o seu metabolismo é um dos alvos de mutação. Os resultados destes ensaios suportam a hipótese da *E.coli* no intestino do ratinho estar a evoluir para o consumo de várias fontes de carbono em simultâneo e em baixas concentrações, por outras palavras estes clones estão a evoluir para serem generalistas num ambiente pobre em nutrientes.

Os resultados aqui apresentados demonstram um grande paralelismo na adaptação da *E.coli* ao intestino do ratinho. A reprodutibilidade da evolução da *E.coli* no ambiente complexo do intestino mostra que a evolução paralela não está restrita aos ambientes de laboratório mas também pode ser comum em ambientes naturais.

**Palavras-Chave:** *Escherichia coli*; Adaptation paralela; Ambiente natural; Microbiota; Generalistas; Evolução molecular; Sequências de Inserção; Estudos de expressão génica.

## Introduction

### 1. Mammalian gut and its inhabitants

The mammalian (with emphasis in human) gut microbiota has become a subject of extensive research in recent years and the knowledge of the resident species and their potential functional capacity is rapidly growing.

The human gut has being described as harboring an estimated $10^{14}$ cells of different species of bacteria and archaea establishing a highly diverse microbial ecosystem(6). This ecosystem is named microbiota and in association with the host has been described as a superorganism. Here the microbiota acts as a virtual organ that provides multiple services to the body(7,8).

The microbiota beyond being huge in numbers, harbors a vast genetic potential estimated to account for roughly 100 times the number of genes present in the human genome (9,10).

The microbiota colonizes not only the gut but virtually every surface of the human body that is exposed to the external environment. Microbes flourish on our skin and in the genitourinary, gastrointestinal, and respiratory tracts(Fig1) (11).

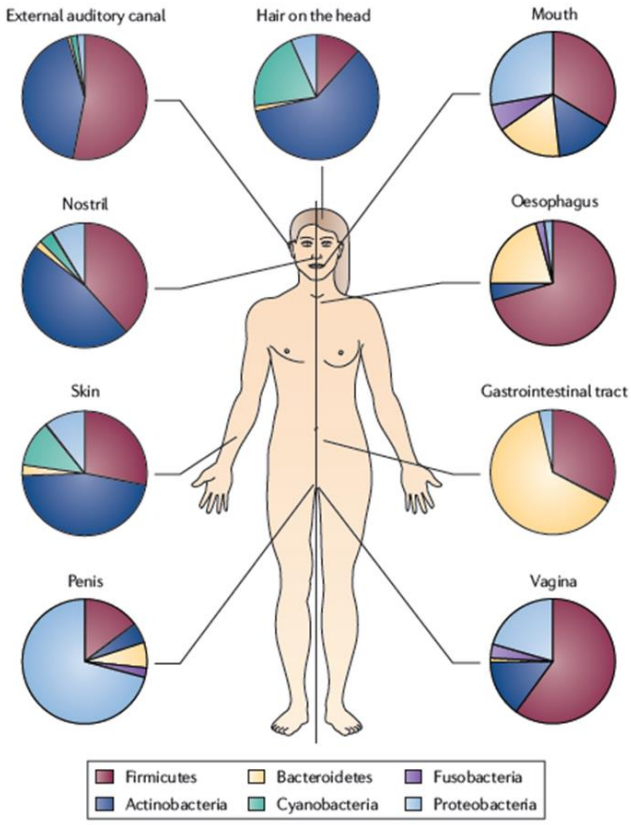


Figure 1 –Microbial community composition at different body locations in a healthy human adapted from Spor et al in (11). The relative abundances of the six dominant bacterial phyla in each of the different body sites: the external auditory canal, the hair on the head , the mouth, the esophagus, the gastrointestinal tract, the vagina, the penis, the skin and the nostril.

By far the most heavily colonized organ is the gastrointestinal tract (GIT). The colon alone is estimated to contain over 70% of all the microbes in the human body (6,12), since the gut is rich in molecules that can be used as nutrients by microbes, making it a preferred site for colonization. The composition of this microbiota has been address by several studies. One example is Arumugam et al (13) that have characterized the phylogenetic variation across several human microbiomes at the phylum and genus level as is demonstrated in Figure 2. They showed that the Firmicutes and Bacteroidetes phyla constitute the vast majority of the dominant human gut microbiota. At the genus level they showed that the dominant were Bacteroides, Faecalibacterium and Bifidobacterium.

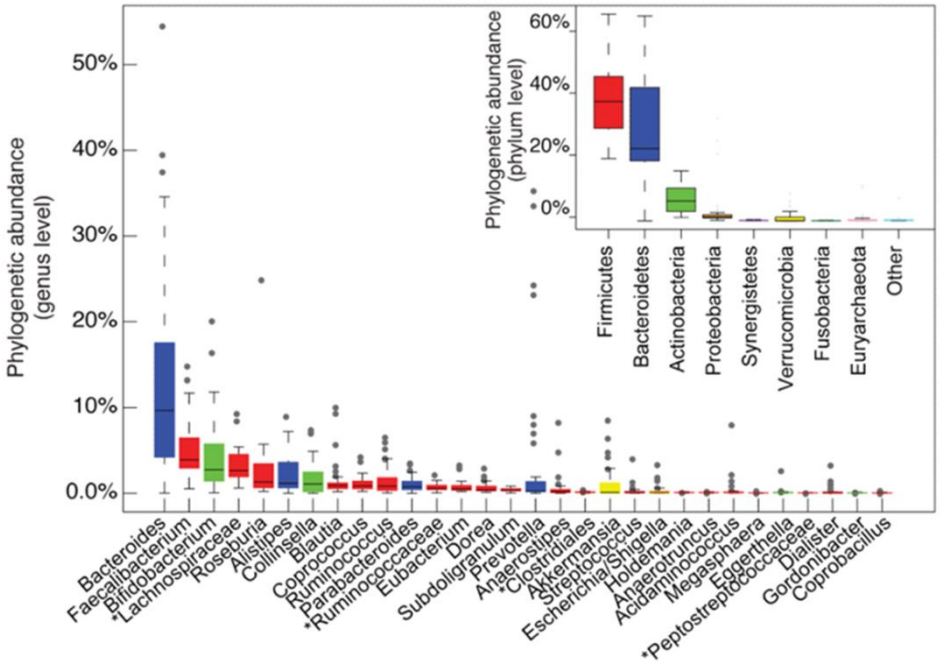


Figure 2 – Phylogenetic profiles of human gut microbiome adapted from Arumugam M. et al in (13). Genera are colored by their respective phylum (see inset for color key).
Inset: phylum abundance box plot. Unclassified genera under a higher rank are marked by asterisks.

The understanding of the complexity of intestinal communities as well as their genetic and metabolic potential is revealed of great importance. In the context of this community diversity means balanced and good microbiota (14).

## 2. *Escherichia coli* in its natural environment: the mammalian gut

The model organism used in this thesis is *Escherichia coli* which was discovered in the late 1800's by Theodor Escherich, a pediatric physician chasing the cause of the fatal intestinal diseases in children. It was first isolated from children fecal samples (15).

Bacteria can live in many different environments and this diversity may be due to independent ecological specializations. *E. coli* is a species that can inhabit a variety of environments, being one of the predominant facultative anaerobes of the intestinal tract but also capable of occupying other secondary habitats (16) *E.coli* can colonize a variety of hosts with distinct body size, diets, morphologies or microbiota, being present in more than 90% of humans(17), 56% of wild animals, 23% in birds and 10% in reptiles (18). Despite being present in approximately 90% of the individuals of our species, the frequency of facultative anaerobes including *E.coli* present in the human microbiota is less than 1% in healthy individuals (1).

*E. coli* usually colonizes the gastrointestinal tract of newborns in the first 40 hours after the birth, though the colonization mode is influenced by the mode of delivery, infant diet, hygiene levels and medication (2,19). Moreover recently it has been found to be the most abundant species in the human placenta (3). Some strains remain present for months to years while others are more transient and remain only for few days (20,21).

*Escherichia coli* belongs to the phylum *Proteobacteria* and class *Gammaproteobacteria*. The order of this bacterium is *Enterobacteriales*, the family is *Enterobacteriacea* and the genus is *Escherichia*. It is a Gram-negative rod-shaped bacterium belonging to the group of facultative anaerobes. Facultative anaerobes are bacteria which prefer to grow in the presence of oxygen, but that can continue to grow without it. It is accustomed to a pH of 7 to 8 and a body temperature of 37Cº.

Despite *E.coli* being one of the best characterized model organisms, it is not only a collection of strains growing in laboratories. In the wild its total population size has been estimated to be approximately $10^{20}$ (4).This bacterium can be classified into major phylogenetic groups and has a high level of genotypic and phenotypic diversity. Most strains of *E.coli* are commensal, but others can adopt the form of a deadly pathogen. With genome analysis from twenty different *E.coli* strains, it was found that the average *E.coli* genome contains 4721 genes, but that only 2000 genes with high homology are shared between all strains. This set of conserved genes was named the core genome (22).

Currently, *E. coli* strains present in human and animal populations are divided into four major recognized phylogenetic subgroup : A, B1, B2 and D (4,23). In humans, considerable variation has been observed in the proportions of *E. coli* subgroups during the past 20 years (4).

*E. coli* pathogenic strains are divided into several pathotypes, such as enterohemorragic, enteropathogenic, enterotoxigenic, enteroinvasive and extraintestinal pathogenic *E.coli* (24,25). Pathogenic strains of *E.coli* can cause urinary tract infections, gastroenteritis, neonatal meningitis, hemorrhagic colitis, multifactorial hemolytic uremic syndrome, peritonitis, mastitis, septicemia and other diseases (26).

The commensal *E.coli* used in this study belongs to the group A. It has a rapid growth rate in lab conditions, well-established genetics and the complete genomic sequence has also been determined. These features make it one of the most well-known model organisms. The relevance of studying *E.coli* comes from the fact of this being the perfect candidate to study the transitions in the interaction between bacterium and the host, which can fluctuate between mutualism, commensalism or opportunistic pathogenesis (4).

### 2.1 Microbiota – Host interactions

In the natural environments, many abiotic as well as biotic factors shape the environment to which microbes adapt.

Colonization and maintenance of bacteria in the mammalian gut can be influenced by many factors. These factors include the host immune system, nutrients, oxygen level, pH, temperature, the presence of antibiotics, diet from the host or even the competition and cooperation between the members of the microbiota.

Normally, the microbiota remains stable for months, and possibly even years (27). However the stability of these microbial communities can be quickly and profoundly altered by common human actions and experiences (28).

Two bacterial phyla, the Firmicutes and the Bacteroidetes, commonly dominate this human gut complex ecosystem(29), as they do in the guts of at least 60 mammalian species (30). These commensal microbes can have many functions in the cooperative process such as training the immune system. Despite the symbiotic nature of this association, the dense communities of bacteria from the microbiota can try opportunistic invasion of the host tissue and cause serious health consequences to the host. There are several examples like commensal *E.coli,* which is present in the microbiota from the gastrointestinal tract, where it normally does not cause disease (8). Nevertheless, even commensal *E. coli* can become an opportunistic pathogen if it breaks hosts physiological barriers, causing urinary tract infections or sepsis (31). Other opportunistic pathogens such as *Candida* sp and *Clostridium* sp also fulfil a beneficial role in the gastrointestinal tract when at a low level, by contributing to the maturation of the immune system; nevertheless, when perturbed the microbiota balance, they could incite disease(32).

The immune system has evolved adaptations to keep the balance in the microbiota and preserve the homeostatic relationship between host and microbiota (33). The microbiota is thought to have other functions like the vitamin production and harvest of otherwise unavailable nutrients sources (34). It is described as even preventing pathogen infection (35). Gut microbiota can provide protection to its host, working as a physical barrier to incoming pathogens by competing for the attachment sites, or consuming the nutrient sources, and producing antimicrobial substances (14).

Gut microbiome can be modulated by dietary changes, antibiotic use, or disease (28), disrupting the balance between the microbiota and the host. Identifying specific disease-associated signatures in the microbiome as well as the factors that alter the microbial populations and gene expression will lead to the development of new products such as prebiotics, probiotics or other drugs to treat these disorders. The World Health Organization has defined probiotics as the "*live microorganisms that can provide benefits to human health when administered in adequate amounts, which confer a beneficial health effect on the host"* (WHO/2001). Another way of shaping a healthy microbiota is through the use of prebiotics. These are composed of oligosaccharides that the host cannot digest and that can have a beneficial effect on health through the selective stimulation of growth and/or activity of specific members of the gut microbiota (36). Live bacteria (probiotics) and prebiotics can be administered together in order to improve their function, and consequently improve the health benefits to the host (37).

For instance, probiotics can modulate the immune system by increasing the production of pathogen-fighting antibodies and other immune cells. They are also able to enhance the function of intestinal cells by stimulating them to produce more mucus and protecting them from dying off. Other studies have shown how probiotics can act as antimicrobial agents by suppressing the growth and invasion of harmful pathogens. From a therapeutic point-of-view, probiotics and prebiotics have been reported to help in the treatment of numerous gastrointestinal disorders (for more information see Vieira et al (37)). However, further tests for potential side effects both in the host and in the bacteria are still necessary.

### 2.2 Microbiota in disease

Nowadays research efforts focused on elucidating the contributions of the gut microbiota to a considerable number of complex diseases. Studies on obesity for example have revealed that diet can affect the gut microbial composition(29) or that microbiome from an obese can have increased capacity to harvest energy from the host diet (38). Several other diseases like allergies, type 1 diabetes, inflammatory bowel disease or even autism are now beginning to be associated with microbiota disruptions ( for more details see Serikov et. al.(14)).

The disruption of the microbiota created by the administration of antibiotics or by a disease has the potential to adversely affect the function of multiple host organ systems for a prolonged period of time (14). Additionally, exposure of the microbial inhabitants of the gut to various antibiotics is likely to result in the acquiring of multiple antibiotic resistances by these commensal bacteria, which can further delay or prevent the return to equilibrium if the same antibiotic regimens are administered repeatedly. Furthermore, it can promote the spread of antibiotic resistance among pathogenic bacteria that will come in contact with the antibiotic-resistant microbiota (14).

Strategies that try to restore the normal gut microbiota have been extensively studied in human and animal models, as these methods represent a valuable tool to treat several disorders. One of these techniques is the fecal microbiota transplantation (FMT) which consists in the transference of a fecal suspension from a healthy person into the gastrointestinal (GI) tract of another person to restore the microbiota and possibly cure a specific disease. Successful examples of applying this strategy have been accumulating with one of the most important being the effectiveness of fecal transplantation to patients with recurrent *Clostridium difficile* infections(39–41). These studies are being performed all over the world in order to understand the capacity of this treatment to cure other diseases.

## 3. Selective Pressures experienced in the gut

Given *E.coli*'s potential to adapt to simple abiotic and complex biotic environments, it is important to understand which type of selective pressures it encounters in its natural environments. The selective pressures in the habitats of commensal strains may coincidentally promote the emergence of virulence factors and antibiotic resistance.

### 3.1 Nutrient availability as selective pressure in bacterial populations

As genetic information contained by the myriad of gut microbes encodes for a far more versatile metabolome than that found in the human genome (42), nutrient availability will be an important feature that can influence their maintenance in the gut. While studying the mechanisms that control bacterial populations in continuous-flow culture models of mouse large intestinal flora Freter et al (43) concluded that the competition for resources is a very important factor, leading him to propose the nutrient-niche hypothesis. This theory postulates that each individual species has a preference for one, or a few nutrients present, creating a nutrient-defined niche occupied by an individual species (43). In other words the gut is a balanced ecosystem in which only one species can occupy one of the numerous intestinal niches presented. As a consequence, the size of the populations of individual species is assumed to be determined by the availability of the preferred nutrient. The relatively low population size of *E.coli* in the intestine indicates that the concentration of its preferred nutrients is low (44,45).

In the large intestine, bacteria grow on nutrients acquired from the mucus layer (44–46). By performing gene expression profiles of *E.coli* MG1655 while growing in mucus, Chang et al (44) identified which genes were up-regulated in this situation. The main targets were genes involved in catabolism of NAG, sialic acid, glucosamine, fucose, ribose, glucuronate, galacturonate, gluconate and maltose. In the same work Chang et al described which subset of carbon sources is mostly likely to be involved in *E.coli'*s colonization and maintenance in the gut. They proposed gluconate as the major carbon source for colonization whereas glucuronate, mannose, fucose and ribose proved to be important for *E.coli'*s maintenance. Therefore they

concluded that colonization of the mouse gut by *E. coli* MG1655 is supported, to various degrees, by at least seven mucus-derived carbon sources. These carbon sources were used by us, to test for adaptive phenotypes of *E.coli*, in in vitro competitions (results-section3).

### 3.1.1 Specialists vs Generalists

An additional level of complexity in evolutionary studies is added when abiotic environmental conditions change over time and space (47–49), which is very likely to be the case of the gut. An interesting example of an *in vitro* experiment where this problem was addressed is the work by Cooper and Lenski (47) where they studied 42 populations of *E. coli* propagated for 2,000 generations either on single carbon sources (glucose, lactose or maltose) or in combination of two carbon sources (glucose and maltose or glucose and lactose) presented together or fluctuating daily. Interestingly, they found that populations adapting to fluctuating environments showed greater fitness variation among populations than in constant environments. Regarding to the specific carbon sources, populations evolved in glucose improved the least, whereas populations evolving in maltose or lactose showed the largest fitness increase, even when compared with the respective mixture with glucose, so the mutants specialize in one carbon sources.

In spatially structured environments mutants can evolve that are better adapted to particular regions and/or particular nutrients. In contrast in continuous nutrient-limited environments, theory predicts the selection of strains that scavenge the limiting resources or more efficiently convert that resources (50). It is recognized that whereas adaptation to pure carbon sources can favor the evolution of specialists, adaptation on mixed carbon sources provides an opportunity for generalists to evolve.

Given this, what can be expected in the complex environment that is the gut? What is the trait that we possibly will observe in there? Will they be specialists or generalists?

While specialists are the best in one specific environment, that is, at metabolizing one single carbon source, generalists can metabolize a variety of them. The drawback of being a specialist is the cost of the adaptation, which could be deleterious in a different environment. On the other hand being a generalist brings the problem of the generalization costs, that means they can grow in several different environments but cannot be the fittest in none of them (51).

### 3.2 Oxygen levels as selective pressure in the mammalian gut

Other type of pressure influencing the microbiota present in the gut is the oxygen level. Oxygen tension is low in the intestine. Using several specific techniques He et. al.(52) was able to map spatial differences in oxygen tension inside the gut. These measurements revealed a marked oxygen gradient from the proximal to the distal GI tract. Their hypothesis to explain this observed oxygen gradient along the GI tract is a combination of several processes. From the swelling of food to the transition to the stomach and then to the small intestine, oxygen levels fall, as oxygen diffuses across the mucosal membrane. On the passage to the colon the presence of bacteria will decrease even more this oxygen tension, being almost none in the distal part of the lumen of the colon.

Anaerobic respiration during evolutionary history use compounds such as fumarate, nitrate or succinate as electron acceptors. After the rise of atmospheric oxygen concentration caused by the metabolism of photosynthetic organisms, aerobic respiration emerged using oxygen as terminal electron acceptor. Since oxygen has a higher redox potential than the electron acceptors used in anaerobic respiration it permits the release of more energy per oxidized molecule. Several types of oxidation-reduction enzymes participate in this electron transport. These include NADH dehydrogenases, flavoproteins, iron-sulfur proteins, quinone, cytochromes in aerobic pathway and reductases as nitrate and fumarate reductases in the anaerobic electron transport pathway (53).

Through the possibility of the presence of oxygen in the intestine other experiments where done in order to understand *E.coli*'s (a facultative anaerobe) metabolism in the intestine. It was observed that *E.coli* lacking the high-affinity cytochrome bd oxidase (used when oxygen tension is low) failed to colonize the mouse intestine. On the other hand the low-affinity cytochrome $bo_3$ oxidase (used when oxygen tension is high) was found not to be necessary for colonization. It was showed that other genes like the *arcA*, for anaerobic respiratory control, *narG* (gene coding for nitrate reductase), or even the *frdA* (gene coding for fumarate reductase) were all necessary for colonization. These observations suggest that *E.coli* is dependent on both microaerobic and anaerobic respiration, leading to the conclusion that it maximizes its growth yield by coupling oxidation of low nutrient concentration to respiration in the intestine (54).

## 4. Evolutionary mechanisms that can underlie the adaptation

Bacterial populations are powerful models to explore the mechanisms of evolution. *E.coli* is part of a massive variety of microorganisms that inhabit the gut. This diversity is one of the strongest selective pressures acting on *E.coli 's* evolution in the gut (55). The ability to constantly adapt is crucial to survive in this complex environment(56).

Although *E.coli* is one of the predominant facultative anaerobes, how it colonizes and adapts growing in the intestine of mammals is still not well understood. In this work we are trying to understand the forces that are shaping its evolution and which is the evolutionary mechanism that plays a central role in such evolution. There are several well-known evolutionary mechanisms that can explain this adaptation. Mechanisms like mutation that are modifications in the gene sequence, which can occur in a myriad of ways: by single nucleotide changes, deletions or insertions, by duplications or translocations, or by the movement of transposable elements, such as insertion sequences (IS).

Usually adaptation depends on the frequency at which new mutations arise in natural populations and their fitness effects, which are of crucial importance for endurance of organisms to the novel environment (57).

Insertion Sequences (IS) are a type of Transposable Elements (TEs) very common in bacterial genomes(58) and with an outstanding ability to mobilize within the host genome (59). These TEs are thought to be major players in the genome organization and evolution and found to have an important role in our experiment. Many times these elements are described as encoding nothing more than their own mobility. They are very common in bacterial genomes. Although most of their transpositions will decrease bacteria's growth rate, they are thought to play a major role in bacteria adaptation processes to specific environments (60).

Other important mechanism is genetic drift. It is characterized as the random sampling of alleles from one generation to the next, (particularly important in small populations). The exchange of genetic information between different strains (recombination) and natural selection are also important in explaining the diversity in populations. In adaptive evolution experiments, selection can drive different evolutionary dynamics. First if a mutation reaches genetic fixation before others become established,(if their rate of appearance is low relative to its fitness advantages and to population size) this leads to a type of dynamics called periodic selection (Figure 3a) (61). However if the rate of beneficial mutations appearance is high (62), it is very unlikely that one beneficial mutation can reach fixation before another appears in another lineage. In asexual populations this means that these two mutations will compete between them, thus decreasing the probability of fixation. This effect is called clonal interference and gives rise to a type of dynamics characterized by soft genetic sweeps, where multiple adaptive alleles at the same locus that can sweep through the population at the same time (5,63)(Figure 3b).

Selection is so strong that even with more than one beneficial mutation appearing; one mutation outcompetes the others and gets fixed in the population. This effect translates into dynamics with hard genetic sweeps (Figure 3c).

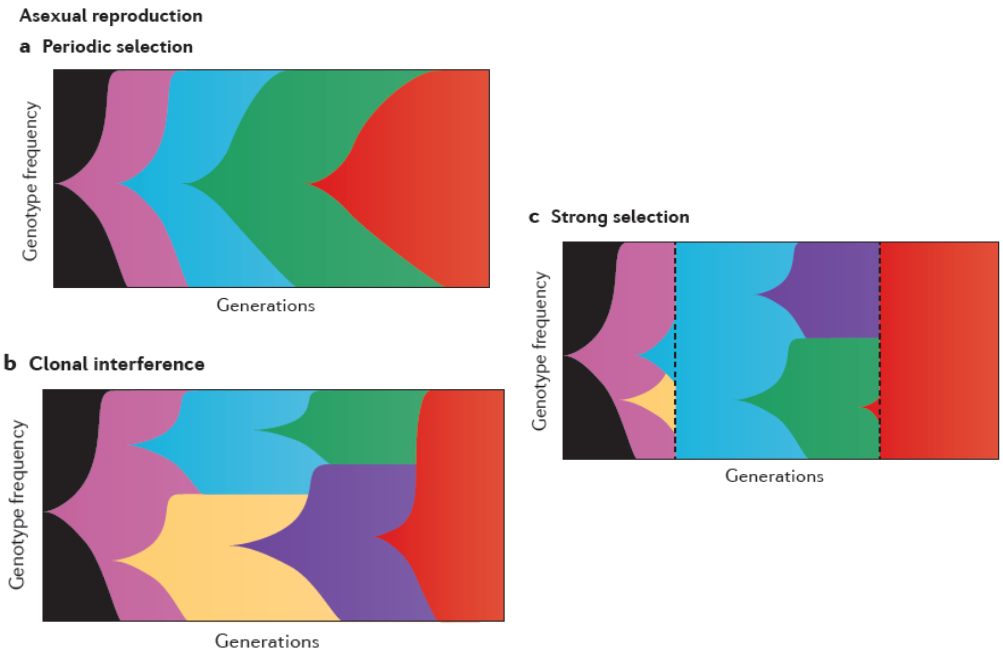


Figure 3 –Genetic dynamics in evolution experiments in asexual organisms adapted from Barrick JE and Lenski RE in (50). a – Periodic selection regime. b- Clonal interference, giving origin to "soft sweep" mutations. c- If strong selection is periodicaly imposed, in ways that may be lethal to most of the population (shown by dashed vertical lines), named of "hard sweeps".

Dynamics can be driven by ecological interactions within and between species like the ones that can be anticipated to occur in the intestine. These interactions can lead to negative frequency dependent selection. Negative frequency-dependent selection translates in fitness advantage of clones when at low frequency but fitness disadvantage when at high frequency. Importantly this will lead to maintenance of genetic diversity (64–66).

The evolutionary mechanisms presented above have been mostly described for evolutionary experiments *in vitro*. It is of upmost importance to understand their importance when in a natural scenario like the gut.

## 5. *In vitro* vs *In vivo* experimental evolution

*In vitro* experiments are really important for evolutionary biology because they provide an enormous level of simplification allowing the investigator to focus in only a specific trait/component. In our case we will test one by one, the potential metabolic pressures faced by *E.coli* inside the mammalian gut. Although many issues can be addressed by in vitro experiments, these are normally too focused in only one trait but in order to address a natural environment with all the interactions biotic and abiotic, in vivo experiments have to be performed and so we did in order to understand evolutionary mechanisms underlying the *E.coli*'s adaptation to the gut and to understand the rate of this process in a natural environment.

One of the most used model organisms to address issues related with the microbiota is the mouse. In this thesis we use streptomycin-treated mice. Streptomycin mimics a typical inflammatory environment in the intestine, showing a decrease in the microbial richness and diversity and being accompanied by a reduction in obligate anaerobic bacteria creating an enabling environment for the outgrowth of facultative anaerobes like *E.coli*. In short, streptomycin opens a previously unavailable niche, which can then be colonized by newly introduced microorganisms such as *E coli* (67,68).

In the present work we study the genetic basis of the adaptation of *E.coli* in two consecutive colonizations of the mouse gut. Fifteen streptomycin-treated mice were "gavaged"[1] in each colonization, with a co-culture of two strains of *E.coli*, each marked by a chromosomal encoded fluorescence and otherwise genetically identical. The strain used in the second colonization was a selected clone from the last time point (isolated from one mouse) of the first colonization. The genetic basis was described by performing Whole Genome Sequencing (WGS) and the effect of the parallel mutations on gene expression, specifically the effect of the insertion of IS elements in regulatory regions was quantified by RT-qPCR. We found that some mutations have a differential effect in the gene expression dependent of the oxygen levels present in the medium. Some genes revealed to be more expressed when in anaerobiosis and others in aerobiosis enlightening the complexity of the oxygen levels inside the gut. We also determined the haplotype diversity during the adaptation to better understand the dynamics of adaptation. The haplotype diversity (from 4 and 3 analyzed populations from first and second colonization respectively) shows a regime of clonal interference. The number of haplotypes emerging in the first colonization was higher than the number in the second colonization. Furthermore we investigate the potential metabolic pressures faced by *E.coli* when inside the gut, growing the evolved clones from both colonizations in different environments comprising single carbon sources and a combination of carbon sources described to be present in the intestine (44).

[1]Direct delivery of a suspension of cells to the stomach by means of a tube

## Results

### 1. Neutral marker dynamics during the first and second steps of adaptation of *Escherichia coli* to the mouse gut

As an adaptive walk is expected to involve more than one adaptive steps, we have studied the adaptive evolution of *E.coli* corresponding to the first and second steps of adaptation to the mouse gut. The first step was obtained by using a commensal lab strain of *E. coli* (K12, MG1655) to colonize fifteen mice and allow it to evolve for 432 generations (5). For this, mice were gavaged with a mixture of equal amounts of two isogenic strains of *E.coli* except for a chromosomally encoded fluorescent protein (CFP or YFP). Next the frequency of the neutral markers was followed by platting the fecal pellets sampled from each mouse every day for 24 days (~432 generations) and counting the number of CFP and YFP colonies (dynamics are shown in Figure 4A).

Following the first period of adaptation we isolated a clone from each mouse and performed WGS. This allowed us to select a clone from the last time point bearing a SNP and a duplication (see table S2.1 – clone 12YFP). Duplications are highly unstable when displaced from the environment where they were selected. Therefore, while genetically manipulating the clone 12YFP to reestablish the polymorphism for the neutral marker, the duplication was completely lost. This was a necessary procedure to obtain the ancestors of the second adaptive step, two isogenic clones (except for the fluorescence marker) differing by a single mutation from the ancestors of the first adaptive period.

The second adaptive step was performed in the same way as the first except for the strains used to colonize the mice (marker dynamics showed in Figure 4B). Some of these dynamics exhibit an initial increase in frequency of one of the neutral markers, followed by replacement by the subpopulation bearing the other marker (example Figure 4A, population 1.9 and Figure 4B, population 2.12), this was interpreted as a typical signature of clonal interference (CI). This replacement is assumed as the markers hitchhike with the successions of beneficial mutations (69). Within these dynamics was also observed cases in which the frequencies of the markers remained stable (example Figure 4A, population 1.6 and Figure 4B, population 2.14) during the 24 days of the adaptive period. The stability of the neutral markers can be expected under different adaptive scenarios: lack of occurrence of beneficial mutations leading to neutral evolution or in contrast, as the occurrence of strong mutations of similar effect in both backgrounds at the same time. This later causes an intense process of CI but looks similar to the neutral dynamics.

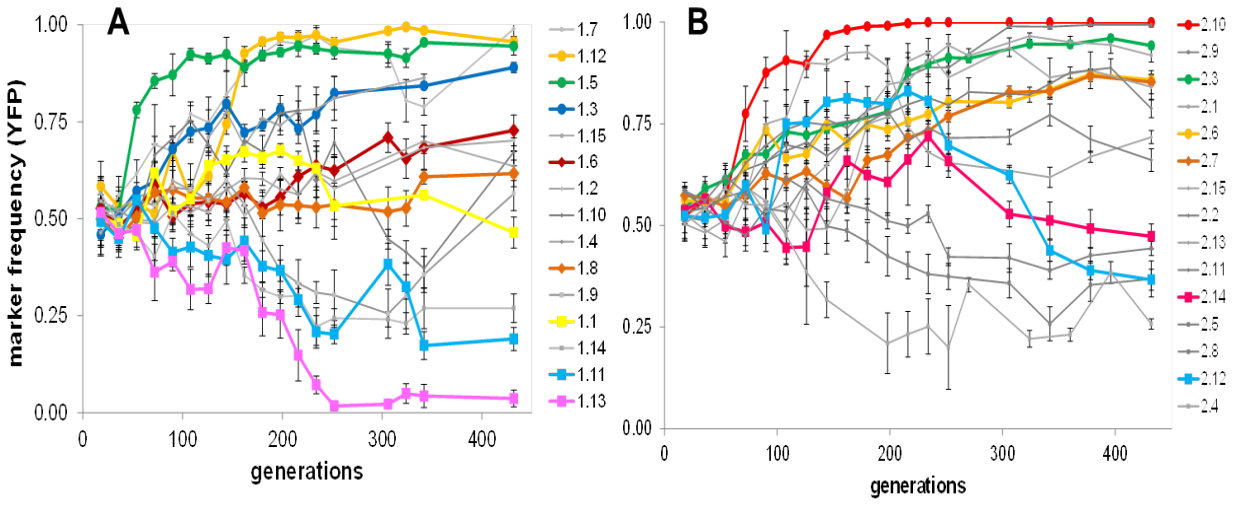


**Figure 4** – Evidence for rapid adaptation **A)** Dynamics of marker frequency during the adaptation of *E.coli* to the mammalian gut in first colonization (1.1 to 1.15). **B)** Dynamics of marker frequency during the adaptation of *E.coli* to the mammalian gut in the second colonization (2.1 to 2.15). Populations shown in colors were used in *in vivo* competitive assays.

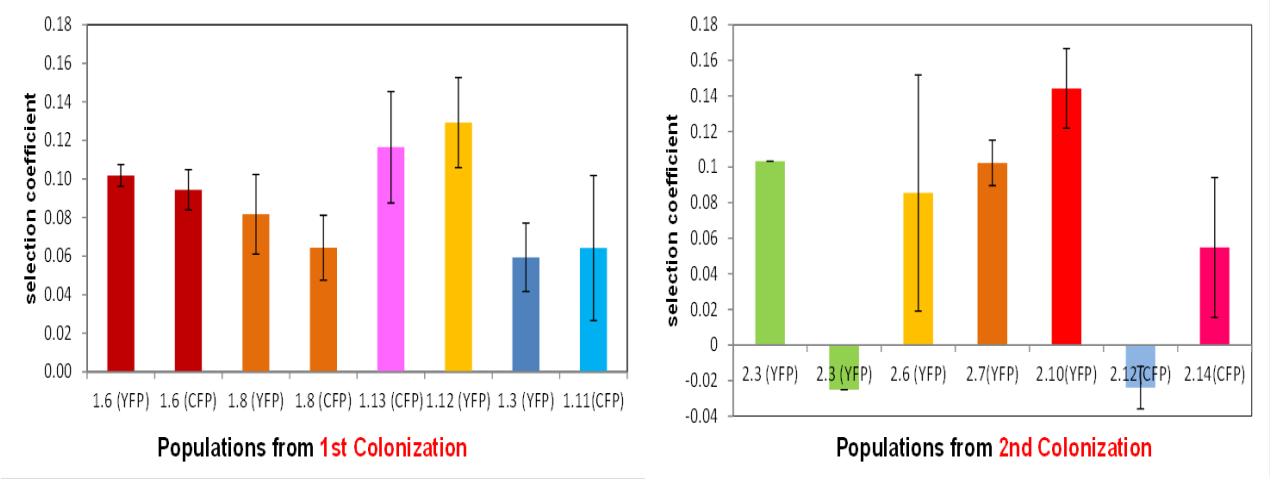


**Figure 5** – Each bar represents the result of a competition between of a mixture of thirty clones (with a given fluorescent marker) isolated from the respective population (indicated below each bar) and the respective ancestor (means selection coefficient ±2 s.e.m, n=3 for the first colonization and n=2, with the exception of population 2.3 that the two replicates are shown separately due to its variability). These results provide direct evidence for adaptation and clonal interference.

In order to distinguish between the two scenarios and to further show that strong adaptive mutations were occurring we performed *in vivo* competitive fitness assays between a mixture of evolved clones (subpopulation labelled with the same neutral marker and isolated from the same mouse) and the respective ancestor clone (either from the first colonization (5) or second colonization (Fig.5)) labelled with the opposite marker. We tested populations where a clear signal of adaptation (that is a rapid deviation from the initial ratio of ~50% of one of the markers) was detected (Fig.5. first colonization – populations 1.12, 1.13, 1.3 and 1.11; second colonization – 2.3, 2.10 and 2.12) and populations where the neutral markers remained polymorphic (Figure 5. first colonization – populations 1.6 and 1.8; second colonization – 2.6, 2.7 and 2.14).

Most clones were found to have increased fitness in comparison with the ancestor (Fig 5). Despite the similar results from both colonizations, two populations from the second colonization showed decreased competitive fitness. Sub-populations from 2.3 and 2.12 showed different fitness effects in competitions performed. Population 2.12 showed a fitness decrease when inside the gut, raising two possibilities. First, the mutations present in this population could only be beneficial in maintenance in the gut but deleterious at initial colonization. Secondly this population can be under frequency dependent selection, where the clones from the background sampled are beneficial if at low frequency but deleterious otherwise. Other difference between first and second colonization is shown by population 2.3. It shows different fitness effects in both independent replicate competitions, showing a fitness increase in one and a decrease in the other. These competitive assays are performed with subpopulations composed by thirty random clones sampled from the populations. These clones potentially have more than one mutation each and could have different genotypes which could explain the inconsistency found in population 2.3. From these results it is possible to understand that we still have adaptation in the second colonization but we cannot state if the rate of this adaptation remains constant in relation to the first colonization.

**2. Genetic basis of adaptation in the first and second steps of adaptation**

At the end of each colonization experiment one clone (of the most represented fluorescent marker) was isolated from the *E.coli* population of each mouse (fourteen clones from the first and fifteen clones from the second step). After this sampling, WGS was performed, in order to characterize genetically the clones from the first (5) and the second step of the adaptive process in vivo (Fig. 6 and Table S2). The mean number of mutations per evolved clone was 2.3 in the first colonization (5) and 1.8 in the second colonization. This analysis showed that ten different mutations were found within the clones from the first colonization and fourteen in the second colonization. But the most striking result from this analysis was the parallelism in the adaptive targets (Fig.6, tableS2). The first-step-mutation inactivated the *gat* operon (involved in the metabolism of galactitol (70)) and was common to all lines. By scoring the frequency of the population that lost the ability to metabolize galactitol along time, we were able to trace back the approximate time of its emergence and the subsequent frequency increase in all lines of the *gat*-negative phenotype (see figure S1). The second-step-mutation appeared in the *gat*-negative background and typically affected one of the following targets: regulatory regions of *dcuB, focA, arcA and yjjP* and in the coding regions of *oppB, radA* and *srlR.*

The *dcuB* gene codes for a transporter, one of three transporters known to be responsible for the uptake of C4-dicarboxylates such as fumarate under anaerobic conditions. Its regulatory region was targeted by the insertion of an insertion sequence (IS) element in three independent clones (three different populations) from the first colonization and in three independent clones from the second one. *focA*, codes for a bidirectional formate transporter and its regulatory region was targeted twice in the first and second colonizations. These IS insertions occurred in the regulatory region of both genes and could either knock-out the gene (abolish its expression), decrease or increase its expression. Knocking-out the gene would narrow the range of substrates that *E.coli* is able to use under anaerobic respiration, whereas increasing its expression could increase the rate of intake of these substrates.

Finally we found four independent clones in the first and three in the second colonization with a SNP in *srlR* the "sorbitol repressor". *srlR* is a DNA-binding transcription factor that inhibits the expression of the *srl* operon involved in transport and utilization of sorbitol.

*arcA* (four clones), a transcriptional dual regulator, is part of a DNA-binding response regulator in two-component regulatory system with *arcB* or *cpxA* product. *arcA* product is a response regulator that represses the genes induced by aerobiosis when in anaerobic condition and also activates some genes induced by anaerobiosis. Its product is phosphorylated by *arcB.* A mutation in this gene could have important pleiotropic effects since without this regulator all the genes induced by aerobiosis would be expressed even when not needed. *radA (*two clones) codes for a recombination protein, and *yjjP* that encodes for a predicted inner membrane protein whose function remains unknown.

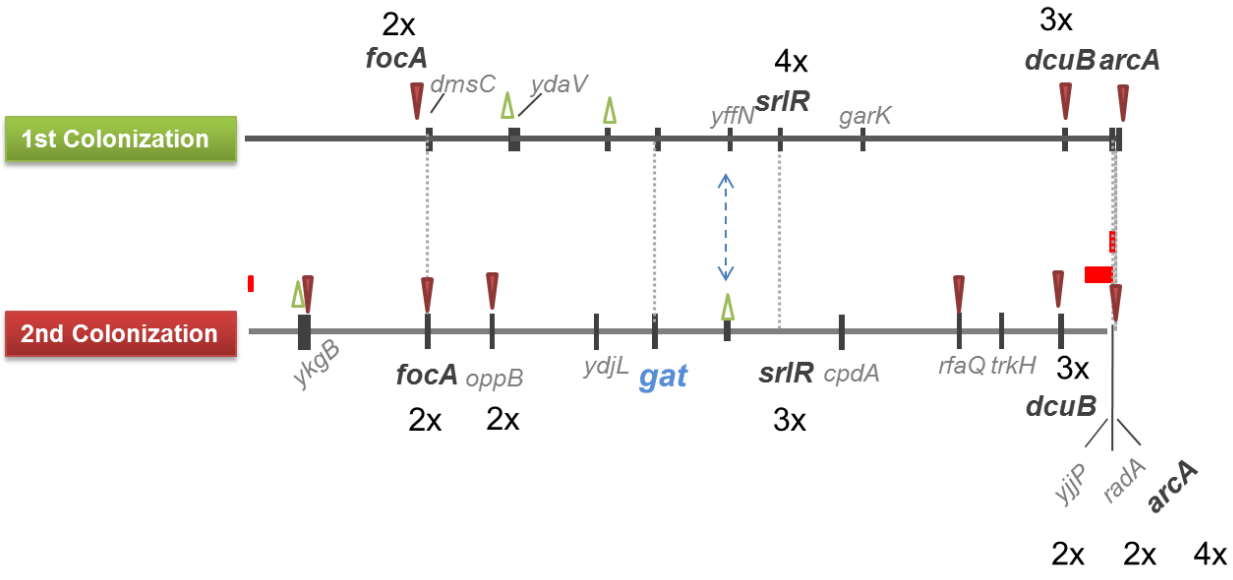


**Figure 6** – Summary of the genetic basis of adaptive mutations in the two colonizations (see table S1 for the complete list of haplotypes). Notice the high level of parallelism observed between clones. Identified mutations in clones isolated from populations from the first and second colonizations (1.1 to 1.14 and 2.1 to 2.15 respectively) represented along the *E.coli*'s chromosome. Inverted red triangles represent IS elements insertions, green triangles represent deletions, red rectangles mark the genomic region duplicated in three clones and the small bars represent SNP's. Genes in bold and dashed lines represent parallel targets for mutation between the first and second colonization. nX represents the number of clones bearing each type of mutation.

**2.1 Quantifying the effect of parallel mutations on gene expression**

Several adaptive mutations were caused by the insertion of an IS element either in the coding or regulatory region of specific genes. IS elements were previously found knocking out the *gat* operon (5). Another case of this IS mutational event occurred in the regulatory region of the *arcA* gene (aerobic respiratory control gene). *E.coli* 's aerobic respiratory control (*arcAB* system) seems to be more important than the aerobic respiration itself when inside the gut (71) so an intriguing question arises regarding the effect of ISs in the regulatory region of certain genes. More generally we wondered about the effects of these mutations, resulting from adaptation to the mammalian gut, at the gene expression level.

To address this question we determined the effect of some of the parallel insertions (Fig.7) by measuring gene expression level by RT-qPCR (see methods). The effects on gene expression were measured in aerobic and anaerobic conditions, since *E.coli* experiences both microaerobic and anaerobic conditions inside the gut (52,54). The medium in which the clones were grown was supplemented with a low concentration of a five carbon sources mixture (gluconate, sorbitol, ribose, glucuronate and mannose). Four of them (gluconate, glucuronate, mannose and ribose) are described as being some of the most important for *E.coli* metabolism in the gut (44). The genes chosen to test were *focA, dcuB, arcA* and *yjjP* because they were targeted in parallel in different populations by IS elements in the regulatory region. *yjjY* was also tested due to the fact that the regulatory region of the *arcA* gene is located close to this gene. *srlA* gene was also studied because it belongs to the metabolic operon of sorbitol and even though not directly targeted by mutation, its expression is under the control of *srlR* (which encodes for the repressor of the sorbitol operon and was one of the main targets for mutation in both colonizations).

The RT-qPCR results showed that the IS element insertion in the *focA*, *dcuB* and *arcA* regulatory regions did not significantly altered the expression of these genes under aerobic condition, but a significant increase expression was observed in anaerobic conditions ($P$=0.006, 0.03 and 0.01 respectively, ttest) (Fig.7). The increase expression of *focA* and *dcuB* genes suggests the possibility of increased efficiency of transport of C4-dicarboxylates like fumarate and others, which have been reported as very important for the anaerobic respiration of *E.coli* inside the mouse gut (54).

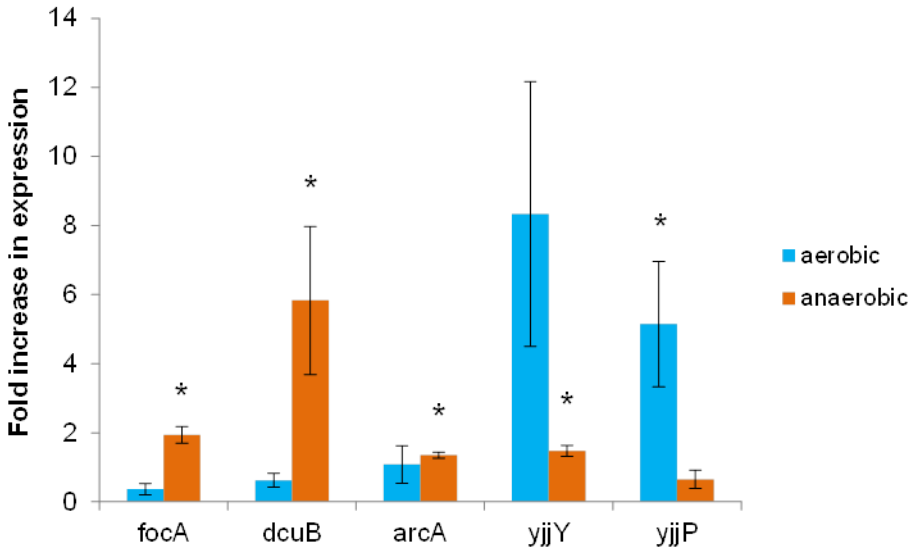


**Figure 7** – Differences in gene expression caused by the insertion of IS elements. These are reported relative to the ancestral strain (expression=1) of the first colonization. Two of the second-step-mutations were omitted from the analysis because the IS element was inserted in the coding region and thus assumed to have inactivated the gene function. Results are shown for aerobic (blue bars) and anaerobic conditions (orange bars). All reactions included three biological replicates for each sample. Data were normalized by the Pfaffl method (83) using the *hfq* housekeeping gene of *E.coli* as a reference. Bars represent the means of three experiments ± standard errors of the mean (SEM). * = *P* value of <0.05 (ttest).

The importance of anaerobic respiration is again revealed by the increase in the expression of *arcA* under these conditions. These results indicate that one of the most powerful selective pressures found by *E.coli* in the gut is probably both availability and variability of the oxygen levels. Other interesting feature that should be taken into account is the fact that these three genes have the IS element inserted in the regulatory region that in contrast to what happened in the *gat* operon, increased gene expression, showing that IS elements not always disrupt genes they can also up regulate the genes(Fig. 7).

The results also showed that *yjjP* gene is being up regulated (*P*=0.02) when in aerobic conditions which suggests an important role for this gene in the aerobic pathways expressed in the gut. *yjjY* expression was measured due to the fact that the mutation in the regulatory region of the *arcA* is near this gene, which could also be affected by it as it was demonstrated in Figure7.

In the context of sorbitol metabolism the results show that *srlA* gene is more expressed in the mutant than in the ancestral strain in both aerobic and anaerobic as shown in Figure 8 (*P* =0.03 and 0.01 respectively, for details of the ttest see materials and methods). The increased expression of the sorbitol operon is consistent with the presence of sorbitol in the gut and with its importance for the success of *Ecoli*'s colonization.

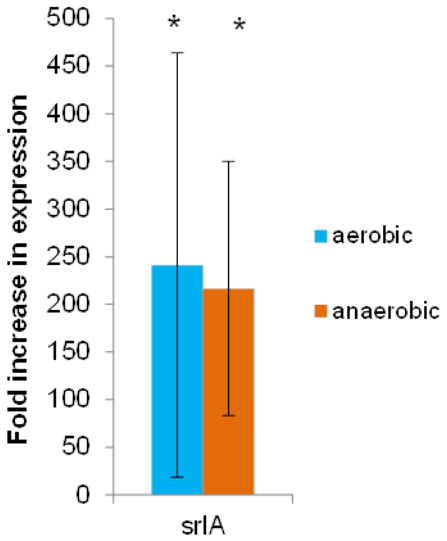


**Figure 8** – Differences in the gene expression of the *srlA* gene.

### 2.2 Haplotype diversity during adaptation

The similarity of the diversity in the dynamics, the results from the *in vivo* competitive assays and the presence of a remarkable parallelism in the mutated genes led us to question whether the rate of adaptation to the mammalian gut is decreasing through time.

In addition to the detected changes in frequency of the neutral markers, we studied polymorphism levels of four of the adapting populations from the first colonization (5) and three from the second colonization in order to better understand the regime of interference. For this we determined the haplotype structure from the first and second colonizations as adaptation proceeded (Fig.9 and 10).

For the first colonization a sample of clones (between 20 and 40 clones per time point, per population) was taken and the haplotypes for the mutations found in the sequenced clones (table S2.1) was determined by target PCR. These included determining the presence of IS insertions in four genes of the *gat* operon (*gatZ*, *gatY*, *gatA* and *gatC*), d*cuB* and *focA*, and SNP's in other two *gat* genes and in *srlR* (5). In order to determine the haplotypes structure in the second colonization we performed whole genome sequencing (see methods) of three of the fifteen populations (2.7, 2.10 and 2.14) in three time points (day11 – gen198, day17 – gen306 and d24 – gen432). We also performed target PCR for the last day of adaptation in populations 2.7 and 2.14, based on the parallel mutations found in the sequenced clones(table S2.2) (2.7 – *arcA* and *dcuB*; 2.14 – *arcA* and *focA*).

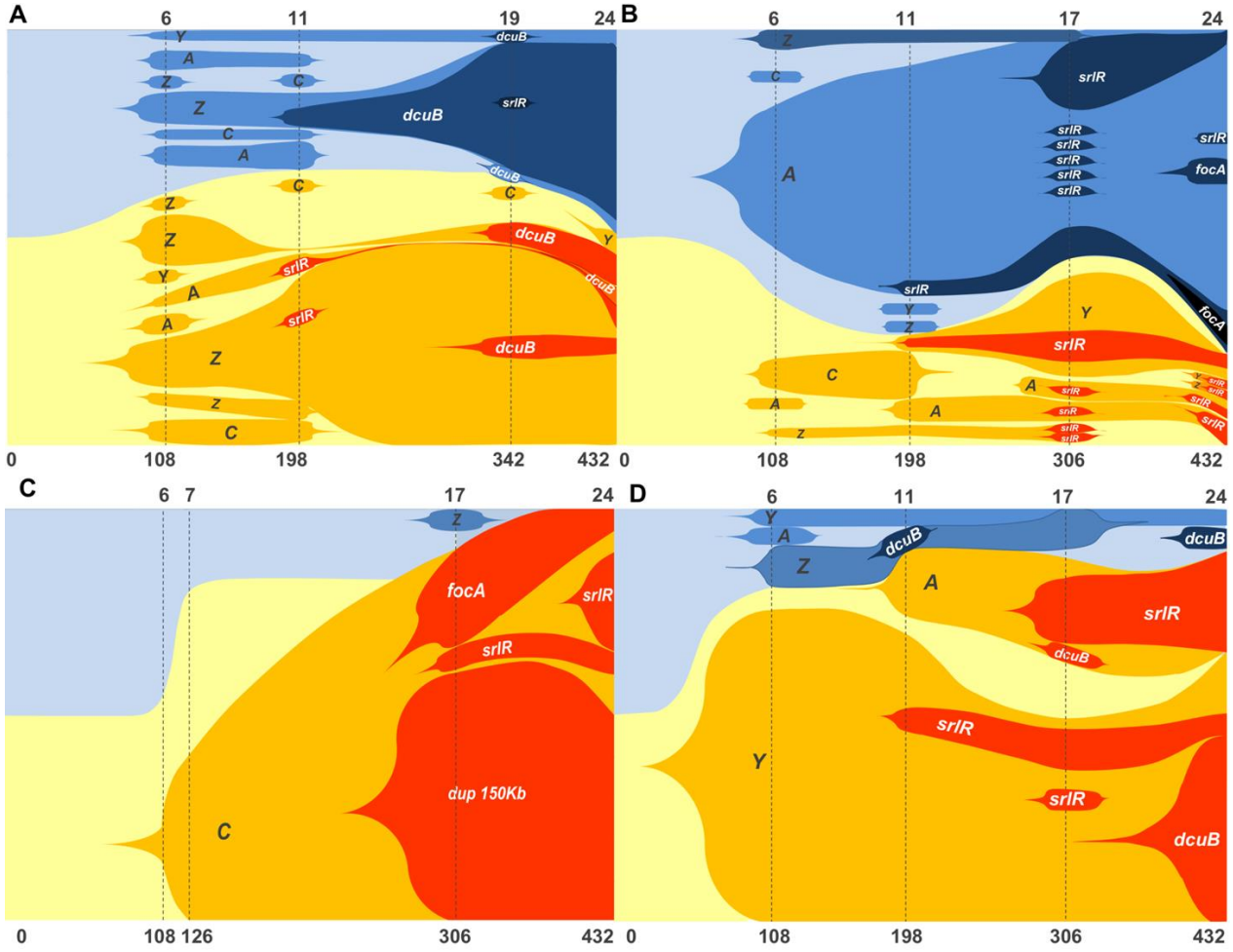


**Figure 9** - Graphic representation of the frequencies of newly generated haplotypes along 24 days (corresponding to 432 generations) of evolution inside the gut during the first colonization. Shaded areas are proportional to the relative abundance of each haplotype. Yellow and blue shaded areas represent the two sub-populations of bacteria labeled either with cfp or yfp alleles. The ancestry relations between haplotypes can be inferred by the accumulation of new mutations in a previously existent genotype. Dash lines mark the time points in days (upper axis) or generations (lower axis) where the sampling took place. For the top two populations 1.1 (A) and 1.11 (B), 40 clones were sampled in each time point. For the bottom two populations 1.12 (C) and 1.5 (D) 20 clones were sampled in each time point. (5)

Figure 9 and 10 shows the dynamics of the haplotypes in the populations studied (first and second colonization respectively).

The genetic analysis from the first colonization showed the appearance of selective sweeps in the sampled populations. Selective sweeps reveal a very intense clonal interference which is caused by strong mutations of similar effects (Fig. 9) (5). Extensive haplotype diversity can be observed in all populations from the first colonization but for a lesser extend in populations 1.12 and 1.5 (Fig.9C and D respectively). On the other hand the haplotype structure from the second colonization populations show a decrease in this interference, where the number of haplotypes was smaller than in the first colonization.

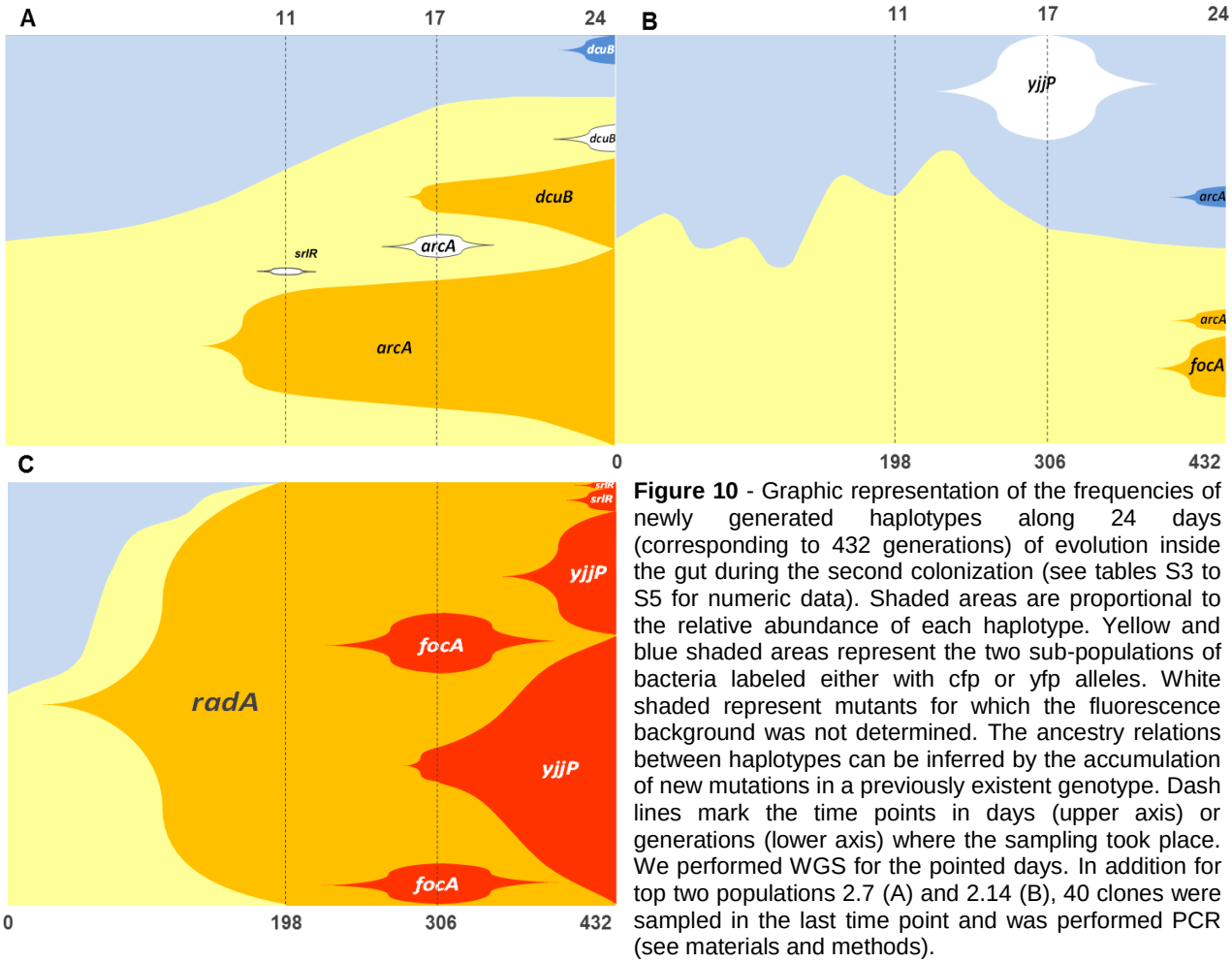


**Figure 10** - Graphic representation of the frequencies of newly generated haplotypes along 24 days (corresponding to 432 generations) of evolution inside the gut during the second colonization (see tables S3 to S5 for numeric data). Shaded areas are proportional to the relative abundance of each haplotype. Yellow and blue shaded areas represent the two sub-populations of bacteria labeled either with cfp or yfp alleles. White shaded represent mutants for which the fluorescence background was not determined. The ancestry relations between haplotypes can be inferred by the accumulation of new mutations in a previously existent genotype. Dash lines mark the time points in days (upper axis) or generations (lower axis) where the sampling took place. We performed WGS for the pointed days. In addition for top two populations 2.7 (A) and 2.14 (B), 40 clones were sampled in the last time point and was performed PCR (see materials and methods).

As Figure 10(A) shows population 2.7 has a haplotype diversity similar to population 1.5 and 1.12 (Fig.9D and C respectively). It has at least 6 haplotypes segregating at the last time point but this haplotypes have only one mutation in contrast with the many double mutants found for the first colonization.

Population 2.14 was the most striking because is one of the cases where the frequency dynamics remained stable. But in contrast to what was found in population 1.11 (Fig.9B) (with similar frequency dynamic) this population is not under the scenario of an intense clonal interference but quite the opposite. On population 1.11 we already observe approximately 10 haplotypes at 198 generations. On the other hand in population 2.14 only one haplotype is seen in this time point (Fig. 10B). Population 2.14 shows less beneficial mutations, pointing to a slowdown of the adaptive process.

In the end when compared with the dynamics of the haplotypes from the first colonization, the second colonization show a significant decrease in clonal interference in both populations 2.7 and 2.14 (Fig. 10A and B, respectively). Besides having less mutational targets, the number of segregating haplotypes with only one mutated gene associated is higher than for haplotypes with more than one gene targeted, as observed in the first colonization. The comparison with the first colonization led us to infer that probably the rate of adaptation of *E.coli* to the mouse intestine is decreasing indeed although *in vivo* competitions could not show it.

We can also observe a distinct pattern of adaptation in population 2.10 showing dynamics of a hard sweep driven by a mutation in the *radA* gene in the YFP background (Fig. 10C) demonstrating a different adaptive dynamic for a second-step mutation that was never found in the first colonization. Mutation in the *radA* gene can bring the possibility for a mutator pattern, and so we performed a fluctuation test (measuring the mutation rate, see materials and methods). The test showed no significant difference between the mutation rates from our *radA* clone (clone 25YFP, isolated from population 2.10) and its ancestral strain. These results are consistent with the result from the sequencing that didn't show the presence of many new mutations, pattern resultant in a mutator strain. In the case of this population the second-step-mutation was probably *radA* and what is observed for the rest of the adaptation period is that *focA, srlR, arcA* and *yjjP* were probably the third-step-mutations. This raises the question of what is the strength of all this mutations, an issue that should deserve future studies.

**3. Testing for potential change in the metabolic traits as a result of adaptation inside the gut**

Given *E.coli's* potential to adapt to simple abiotic and complex biotic environments it is important to understand which type of selective pressures it may encounter in its natural environment.

The mammalian gut is one of *E.coli'*s natural environments in which many factors such as the host immune system, oxygen level, pH, temperature, use of antibiotics, diet from the host or even the competition and cooperation between the microbiota members, can influence its colonization ability. In this section we investigate several metabolic selective pressures such as: the presence of five carbon sources in low concentration either in combination or alone.

We performed *in vitro* competitive assays (see materials and methods) in minimal medium with a mixture of five carbon sources (gluconate, ribose, glucuronate, mannose and sorbitol) described as being some of the most important for *E.coli´*s metabolism in the gut (the first four) and sorbitol, due to the fact that its metabolism was one of the most targeted by mutation during the adaptation to the gut. This mixture was used at two different concentrations. The first mixture had the concentration of 0.01% of each carbon source, which amounts to a final concentration of 0.05% (low concentration) and a second mixture where each carbon source was added in a concentration of 0.1% ( a total of 0.5%, high concentration). Both low and high concentrations were used to address the question whether the mutants isolated from the mouse gut were adapted to grow in a poor environment, or if they were only adapted to the most important carbon sources present in there. These *in vitro* competitive assays give us the relative fitness of the mutants against the ancestor, presented as selection coefficient. This fitness can indicate the strength of selection of particular mutations and can indicate whether these fitness effects differ between different genotypes. The clones tested in these fitness assays were the twenty-nine clones isolated in the last time point of both colonizations (results are presented in table 1).

Table1 shows that almost all clones were fitter growing in the mixture of the five carbon sources than the ancestor. Moreover we can observe that these evolved clones have a higher advantage when growing at a low nutrient environment than with a high concentration of carbon source (P< 0.001, ttest). Therefore these results are compatible with the gut being composed of a multitude of different carbon sources at low concentration. These observations could indicate that the evolved clones are adapting to be generalists in a poor-nutrient environment.

Furthermore competitions were performed in each carbon source separately (in a low concentration) in order to address the issue whether the different clones would have a preferred carbon source. As showed in table 1, in overall, clones have preferred carbon source, having a higher advantage when growing in it. But this preference is not directly connected with a specific genotype, with the exception of the *srlR* (sorbitol metabolism repressor gene) mutants. Clones with the *srlR* mutation have a significant higher advantage when growing with sorbitol as the only carbon source then the others (*P*= 0.004, ttest). This advantage is due to the fact of a probable deficient activity performed by this repressor, as supported by an increased expression of *srlA*.

**Table 1** – Fitness effects of adaptive mutants. This table shows the selective coefficient of each clone when using a mixture of five carbon sources (low and high concentration) or single carbon sources (low concentration). The carbon sources used were gluconate, glucuronate, sorbitol, mannose and ribose, carbon sources described as important for *E.coli* when inside the gut (38). This table represents the results of the in vitro competitions performed in aerobic conditions. Each clone is represented by its genotype and it's organized by the most targeted mutations. In green are represented significant increases in fitness relative to the ancestral strain (dark-light green = major-smaller increase. Red are represents the fitness decreases and in yellow no difference in fitnessin comparison to the ancestral clone. Differences were scored as significant in comparison with the ancestor taking into account the standard errors of each measurement.

| genotype | mix0.05 | mix0.5 | mannose | gluconate | glucuronate | sorbitol | ribose |
|---|---|---|---|---|---|---|---|
| *gatC arcA* | 0.10 | 0.08 | 0.14 | 0.07 | 0.30 | 0.15 | 0.06 |
| *gatC arcA* | 0.11 | 0.05 | 0.17 | 0.17 | 0.14 | 0.01 | -0.11 |
| *gatC arcA(SNP)* | 0.02 | -0.01 | 0.12 | 0.08 | 0.01 | -0.08 | -0.07 |
| *gatC arcA cpdA* | 0.07 | 0.01 | 0.16 | 0.06 | 0.07 | -0.01 | 0.25 |
| *gatC yjjP arcA* | 0.08 | 0.08 | 0.13 | 0.07 | 0.13 | 0.11 | 0.12 |
| *gatY yjjP yffN* | 0.11 | 0.07 | 0.23 | 0.10 | 0.25 | 0.09 | 0.40 |
| *gatC yjjP radA del11kb* | 0.07 | 0.07 | 0.15 | 0.18 | 0.24 | 0.14 | 0.56 |
| *gatC radA* | 0.08 | 0.05 | -0.05 | 0.09 | 0.10 | 0.17 | -0.13 |
| *gatcC dcuB* | 0.06 | 0.06 | 0.12 | 0.05 | 0.15 | 0.15 | 0.18 |
| *gatcC dcuB* | 0.09 | 0.05 | 0.11 | 0.19 | 0.25 | 0.05 | 0.52 |
| *gatZ dcuB* | 0.10 | 0.09 | 0.22 | 0.05 | 0.23 | 0.11 | 0.32 |
| *gatC dcuB* | 0.11 | 0.05 | 0.19 | 0.11 | 0.27 | 0.05 | 0.29 |
| *gatZ dcuB* | 0.10 | 0.05 | 0.18 | 0.14 | 0.17 | 0.09 | 0.20 |
| *gatC dcuB ydjL* | 0.07 | 0.07 | 0.10 | 0.20 | 0.25 | 0.12 | 0.63 |
| *gatC dup150kb* | 0.12 | 0.08 | 0.23 | 0.14 | 0.24 | 0.09 | -0.02 |
| *gatC dup155kb* | 0.13 | 0.10 | 0.15 | 0.26 | 0.37 | 0.13 | 0.60 |
| *gatA srlR* | 0.13 | 0.08 | 0.03 | 0.05 | 0.16 | 0.20 | 0.26 |
| *gatA srlR* | 0.13 | 0.09 | 0.13 | 0.06 | 0.19 | 0.19 | 0.19 |
| *gatA srlR* | 0.09 | 0.05 | 0.02 | 0.14 | 0.18 | 0.17 | 0.08 |
| *gatC srlR oppB* | 0.07 | 0.02 | 0.11 | 0.07 | 0.10 | 0.06 | -0.16 |
| *gatC srlR oppB rfaQ* | 0.11 | 0.07 | 0.00 | 0.24 | 0.37 | 0.21 | 0.20 |
| *gatY srlR dmsC* | 0.10 | 0.06 | -0.03 | 0.15 | 0.18 | 0.17 | 0.08 |
| *gatC focA srlR* | 0.08 | 0.06 | -0.08 | 0.14 | 0.31 | 0.30 | 0.05 |
| *gatZ focA del5kb* | 0.10 | 0.05 | 0.18 | 0.15 | 0.20 | 0.13 | 0.15 |
| *gatY focA ydaV del5kb* | 0.08 | 0.09 | -0.56 | 0.13 | 0.34 | 0.05 | 0.19 |
| *gatC focA trkH* | -0.01 | 0.00 | 0.02 | -0.14 | -0.11 | -0.15 | -0.32 |
| *gatA garK* | 0.10 | 0.09 | 0.24 | 0.13 | 0.29 | 0.08 | 0.06 |
| *gatC ykgB del7kb* | 0.10 | 0.05 | 0.12 | 0.19 | 0.21 | 0.09 | 0.45 |
| *gatZ* | 0.10 | 0.02 | 0.32 | 0.09 | 0.16 | 0.21 | 0.17 |

Legend:

| | |
|---|---|
| | increase |
| | neutral |
| | decrease |

One other important aspect to note is the carbon sources in which more clones have advantage, which are the gluconate and glucuronate. Gluconate is the only in which all the clones have an advantage relative to the ancestor, with the exception of one clone. Supporting the previous observation (44) that gluconate is probably one of the major sources for *E.coli* in the gut. On the other hand ribose was the carbon source in which the fitness effects were more variable.

Other peculiarity that must be noticed in table1 is the fact that the only single mutant (with one mutation in the *gatZ*) has already a great advantage in either the mixture or the single carbon sources, suggesting that the first step of adaptation is crucial to deal with the metabolic pressure faced by *E.coli* inside the gut. Several of the second-step mutations are targeting genes with relative importance in respiratory pathways and some of them have demonstrated to have an effect in the gene expression when in anaerobic conditions. As *E.coli* is known to be dependent on both microaerobic and anaerobic respiration inside the gut (54,71) and because some of the mutations seem to have an effect dependent on the presence of oxygen, in the future we will perform the same competitive assays under anaerobic conditions.

This section allowed us to study the metabolic pressure alone demonstrating that in vitro experiments are an important supplement in order to understand the organism's behavior in its natural environment.

## Discussion

Previously we have demonstrated that *E.coli* can adapt very fast to the intestine of streptomycin-treated mice (5). This first step of adaptation was dominated by soft sweeps, adaptive mutations of large effect and an intense regime of clonal interference where haplotypes carrying either one or more beneficial mutations were competing for fixation. Among the mutations arising during this process a high degree of parallelism was observed not only in the first step, inactivation of the *gat* operon, but also among the targets for the second step (*dcuB, focA, arcA, yjjP, srlR, oppB or radA*). This kind of parallelism already demonstrated in other *in vitro* evolution experiments(64,65), was difficult to anticipate in such a complex environment as the mammalian gut. Here the bacterial community is thought to be under multiple selective pressures, ranging from direct competition with other bacteria, to the action of immune system, the oxygen level or the diet composition.

## 1. Adaptive mutations, a major role for IS elements

High parallelism was found between mutations arising in the adaptation of *E.coli* to the mouse gut. The first gene targeted for mutation was one of the *gat* operon, which was inactivated in all populations (~80% of these events were caused by IS insertions(5)). This operon allows for the galactitol metabolism (70). In our study we use a derived strain of *E.coli* MG1655 which is a commensal *K12 E.coli*. This Wild-type *E.coli K12* has one IS inserted in the *gatR* that disrupts the repressor of the galactitol operon, thus rendering the operon with a constitutive expression. The first hypothesis for this first step of adaptation is the toxicity of the galactitol to our strain since we have found that it had an inhibitory effect on the ancestral strain when grown in minimal medium with other carbon source (5). Other hypothesis was the principle of selection against the production of unneeded proteins (72) which suggest the fitness cost of the expression is proportional to the amount of protein produced, being the production of this useless protein a major energetic cost. The observed benefit that the *gatZ* mutation alone was able to confer in any of the poor nutritional media tested supports this idea (Table 1). Another example was the observation by Stoebel et al (72) that the cost in fitness, when in lactose absence, to *E.coli* strains that are constitutively expressing the lac operon is associated with the actions of transcription and/or translation. The presence of galactitol in the streptomycin-treated gut is yet unkown. If absent this could drive to the hypothesis that the constitutive expression of the *gat* operon has a cost to the cell which would be abolished by the disruption of the operon.
A similar situation was observed by Zhong et al (73) when studying the evolutionary genomics of ecological specialization of an Ecoli K-12 strain which also has an IS inserted in gatR that disrupts the repressor of the galactitol operon. They found that during the specialization process to limiting-nutrient environment one IS was transposed into yegW gene deleting the entire gat operon (73).

Despite the fact that this first step was mainly due to the insertion of ISs, these usually happened in the coding regions being in agreement with the effect of inactivation of the operon and the gat- phenotype. A high rate of insertion of ISs continued in the second step but unlike the first step these targeted mainly the regulatory regions of specific genes. The result of these insertions was difficult to anticipate, since they can cause: abolishing of gene expression, modification of the level or timing of expression or even activate genes previously inactive (some ISs are known to carry promoting sequences)(58). Hence we measured the level of expression of the genes targeted by the IS elements in the regulatory regions (*dcuB*, *focA*, *arcA* and *yjjP*) in aerobic and anaerobic conditions.

One of the second mutations arising is in the regulatory region of the *focA* gene. It is a bidirectional formate transporter with an important role in regulating intracellular formate levels during anaerobic respiration. This gene is extremely important to the maintenance of the pH inside the cells (74). Studies of the formate levels in fermenting *E.coli* cultures indicate that

initially, formate is exported out of the cell in order to prevent acidification of the cytoplasm. However, if the pH of the culture medium drops below about 6.8, formate is re-imported and *focA* has also been implicated in this process (74).

*dcuB* is other gene targeted by an IS in its regulatory region that is important when in anaerobic conditions. This gene is responsible for the uptake of C4-dicarboxylates such as fumarate under anaerobic conditions.

The presence of the IS in the regulatory region of these genes (*dcuB* and *focA*) could cause several consequences to the cell. They could knockout the gene (as in the *gat* genes), or affect its regulation.

The results from the expression level of both *dcuB* and *focA* gene showed that possibly the IS had no effect when in aerobic conditions but triggered an increase in its expression relative to the ancestral strain, when in anaerobic conditions ($P$=0.006 and 0.03 respectively, ttest).

It was already shown that the use of cytochrome bd oxidase as electrons acceptor is more advantageous to *E.coli* when inside the gut. But it was also shown that *E.coli* uses nitrate and fumarate reductase when performing anaerobic respiration. It was found that, when colonizing the intestine, nitrate reductase mutants outcompete the fumarate reductase ones, indicating that fumarate is the more important anaerobic electron acceptor in the intestine because nitrate is limiting (71). So taking into account these observations and our results we can hypothesize that C4-dicarboxylates like fumarate and other anaerobic respiration substrates are present and are important in the anaerobic respirations inside the gut. Our strain appears thus to be adapting by becoming more efficient at transporting these substrates.

*arcA* gene is a transcriptional dual regulator, part of a DNA-binding response regulator in two-component regulatory system with *arcB* or *cpxA* product. *ArcA* product is a response regulator that represses genes induced by aerobiosis when in anaerobic condition and also activates some genes induced by anaerobiosis. An IS element was inserted in the regulatory region of this gene in several populations during the adaptation process. This gene is regulated by other important regulator gene that is the *fnr. Fnr* is the primary transcriptional regulator that mediates the transition from aerobic to anaerobic growth through the regulation of hundreds of genes. Generally, this protein activates genes involved in anaerobic metabolism and represses genes involved in aerobic metabolism (75). The derived *E.coli* strain that was used in this thesis has a deletion in the site of the genome that includes the *fnr* gene. The knock-out of this gene causes down regulation of the *arcA* gene (76) this fact can cause difficulties in *E.coli* colonization and maintenance inside the gut. Marzan et al (76) observed that although *arcA* mutants can colonize when fed alone to the mice, they could not compete with their respective wild type and were eliminated from mice within 3 days (54). The RT-qPCR analysis for this gene showed an increase ($P$=0.01, ttest) expression in anaerobic condition suggesting that appropriate regulation of aerobic respiratory genes is necessary for *E.coli* survival inside the gut. These facts suggest that selection for increase expression of *arcA* might be related with the absence of the important regulator *fnr* and/or to the low oxygen concentration in the gut.

The mutation in the regulatory region of the *dcuB, focA* and *arcA* genes is part of the second step of adaptation, this could suggest that the metabolic optimization is occurring first being the more important selective pressure. But then and because of the oxygen variability in this environment (52), where *E.coli* can encounter microaerobic and anaerobic conditions, becomes important to have respiratory flexibility (better transport of anaerobic respiration substrates and aerobic respiration regulation).
Other peculiarity showed by our results is that the mutations of the regulatory region of *dcuB*, *focA* and *arcA* were never found together (in all combinations *dcuB* and *focA*, or *dcuB* and *arcA* or *focA* and *arcA*). This observation could reveal a process of negative epistasis between them. Negative epistasis happens when the effect of a beneficial mutation is bigger than the combined effect of two of these beneficial mutations. In the case of beneficial mutations this significantly decreases the probablility of finding the double mutant. An example of negative epistasis between beneficial mutations is described in Chou et al (77). They studied the combinations of beneficial mutations that arose in a lineage during rapid adaptation of a bacterium. Results showed that the selective benefit for three of the four beneficial mutations consistently decreased when introduced upon more fit backgrounds (77).

The mutation in the *srlR* gene is possibility due to the metabolic pressures faced inside the gut. The *srlR* gene is the repressor of the metabolism of sorbitol. The mutations targeting this gene were mainly SNP's some of them causing stop codons (5), and a few small indels, suggesting that this gene is probably being severed. The RT-qPCR results on the *srlA* gene (sorbitol PTS permease) showed an increase in expression relative to the ancestral strain in aerobic and anaerobic conditions ($P$= 0.02 and 0.001 respectively, ttest) suggesting that the activity of the repressor (*srlR*) was severely affected. The mutants with the constitutive expression of the sorbitol operon will have an advantage when facing an environment with the presence of sorbitol which can probably be the case of the gut.

The striking observation that the same IS element is targeting the same gene in different bacteria that are evolving in different mice questions the use of these elements as neutral markers of evolution. In certain cases, the localization of different specific IS elements at defined places in the chromosome is considered sufficiently stable to allow them to be used as markers in restriction fragment length polymorphism studies for species typing and for epidemiological purposes (e.g., IS*6100* in *Mycobacterium tuberculosis* (78), IS*200* in *Salmonella* (79) and IS*1004* in *Vibrio cholerae* (80). In other words epidemic studies use ISs as neutral markers to follow pathogenic strains in an outbreak. But this technique may mislead the results. One of the problems is the IS hotspots that exist in *E.coli* genome, which are places in the genome that are more targeted by these transposable elements. Other issue is showed by our results, where the same IS is targeting the same gene but in different bacteria that are evolving independently in the gut of different mice, suggesting that ISs are not the best neutral marker to take into account for epidemiological purposes.

## 2. Selective pressures inside the gut

### 2.1 Oxygen levels

Oxygen level is one of the selective pressures influencing adaptation of *E.coli* to the gut. Some studies revealed that there is a marked oxygen gradient from the proximal to the distal GI tract (52). These observations led to the hypothesis that *E.coli* has being spread throughout the gut being a subpopulation exposed to an environment with more oxygen and other in a more anaerobic one. When in spatially structured environments mutants can be selected to better adapt to particular regions or to better colonize microhabitats. By contrast in continuous nutrient-limited environments, theory predicts they will be selected to scavenge the limiting resource or more efficiently convert that resource (50).

Our results of gene expression showed that at least three of the second step mutations occurring in the adaptation of *E.coli* to the mouse gut are related to the respiration, suggesting that these clones are adapting to the oxygen variability in this environment.

### 2.2 Metabolic Pressure

Specialists and Generalists

Selection in nutrient limited environments and in spatially structured environments can have a huge importance in microbiota balance (50).

There are at least three mechanisms that can origin specialists (81). Clones can accumulate neutral or beneficial mutations, which are deleterious in another environment (mutation accumulation or antagonistic pleiotropy respectively) or they can accumulate beneficial mutations which are neutral in another environment (independent specialization). But only populations growing on single carbon sources can specialize by this three mechanisms, those who grow in mixed carbon sources (as the environment faced by *E.coli* inside the gut) cannot specialize by mutation accumulation because any mutation that have a neutral effect in one carbon source but a deleterious one for a second carbon source will be eliminated by selection. In the other hand the process of independent specialization can drive to the appearance of generalists (51). Our results suggest that the clones that are adapting inside the mouse gut are evolving to be generalists scavenging for nutrients.

In addition gluconate was the only environment where all clones (except one) had a fitness increase relative to the ancestral, supporting the observations already described (44) that gluconate is one of the more important carbon sources for *E.coli* in the gut. On the other hand ribose was the carbon source in which the fitness effects were more variable. It was already observed that the metabolism of ribose both in vitro and in vivo can sometimes be regulated by the presence of other carbon source also important for *E.coli* inside the gut, fucose (44,82). Taking in account this information it is possible that some of the isolated clones evolved with the presence of the two carbon sources and so are adapted to the regulation set by fucose to the metabolism of itself and ribose, which could explain that variability within the clones.

### 3. Evidence for continuous adaptation of *E.coli* in the mouse gut

A common result from in vitro evolution experiments is that the rate of adaptation tends to slow down over time, in other words the fitness increases tend to decelerate over time. An example is the long-term experiment where twelve *E.coli* populations were evolved in glucose and the fitness was measured during the adaptation process. The average fitness gain in the first 5000 generations revealed to be approximately tenfold greater than the one between 15000 and 20000 generations (83).

This can be due to the diminishing of the number of beneficial mutations or the fact that its effect is smaller. Actually this is observed when the first step of adaptation is bigger than the second over the adaptive walk and then the second being bigger than the third consecutively. So it would be plausible that this would happen in the adaptation of *E.coli* to the mouse gut.

With further analysis of the haplotype diversity over time from the second colonization was observed that although soft sweeps were described they were in a small number. In other words the number of haplotypes with different beneficial mutations competing with each other is lower, suggesting the presence of clonal interference but with a smaller strength.

The process of adaptation can be quite complex as described in Good et al (84), who modelled clonal interference. Populations with a rate of mutation very small present a lower number of beneficial mutations enabling the existence of only one mutant in the population at a true classic periodic selection regime. On the other hand in large populations like the one from this study, a much more complicated situation occurs in those distinct mutations segregate simultaneously (84), many of them competing for fixation. These dynamics can create a complex process of interference, which was observed during the first colonization of *E.coli* to the mouse gut (5). This interference is still occurring in the second colonization but possibly with a smaller strength due to the fact that the number of beneficial mutations observed is smaller than in the first colonization.

## Conclusions

In conclusion we demonstrate here that *E.coli* MG1655 continues to adapt in a second colonization to the gut of streptomycin-treated mice but with differences in the target genes and in the adaptive dynamics. The second-step mutations appearing in the second colonization targeted not only *dcuB*, *focA* and *srlR* that were already described in the first as other new targets *radA*, *arcA*, *oppB* and *yjjP*. This second colonization analysis has also demonstrated the presence of soft-sweeps but with inferior clonal interference than the one showed in the first colonization. For example 33% of the sequenced clones from the second colonization carry only one beneficial mutation, in contrast to the first colonization where only one out of 14 clones was a single mutant.

We demonstrate here that IS elements play an important role in *E.coli's* adaptation to this complex environment. Bacterial insertion sequences were initially identified during studies of model genetic systems by their capacity to generate mutations as a result of their translocation. Several studies have revealed an important role for these mobile elements in the genome of bacteria. For example they were found to be associated with many pathogenic and virulence functions, the participation in chromosome rearrangements and in plasmid integration (for more information see Mahillon J. 1998(58)). The evolutionary role of these transposable elements is still controversial.

Regarding the gut environment our results suggest that oxygen level and the metabolic pressures are an important stimulus in the adaptation.

Our results demonstrate the remarkable parallelism in the adaptation of *E.coli* to the mouse gut. Therefore, highly complex, the repeatability of evolution in the gut environment shows that parallel evolution is not restricted to idiosyncratic laboratory environments and may be very common in nature.

## Materials and Methods

### Ethics statement

All experiments involving animals were approved by the Institutional Ethics Committee at the Instituto Gulbenkian de Ciência (project nr. A009/2010 with approval date 2010/10/15), following the Portuguese legislation (PORT 1005/92), which complies with the European Directive 86/609/EEC of the European Council.

### Bacterial Strains

All strains used were derived from MG1655, a K12 commensal strain of *Escherichia coli* .

In the first colonization experiment (used in this experiment) the fluorescent strains MG1655-YFP and CFP (MG1655, *galK::YFP/CFP* $amp^R$ , $str^R$ *(rpsl150), ΔlacIZYA*) were used for colonization. The second colonization started with a descendant from the first colonization that differs from the first ancestral by one mutation (Ins (1bp) *gatC*).

### Fluorescent marker dynamics during mouse colonization

To study *E. coli* adaptation to the gut we used a streptomycin-treated mouse colonization model (Conway et al 2004). Briefly, 6- to 8-week old C57BL/6 male mice raised in specific pathogen free (SPF) conditions were given autoclaved drinking water containing streptomycin (5g/L) for one day. After 4 hours of starvation for water and food the animals were gavaged with 100 µl of a suspension of $10^8$ colony forming units (CFUs) of a mixture of YFP- and CFP-labeled bacteria (ratio 1:1) grown at 37°C in brain heart infusion medium to $OD_{600}$ of 2. After the gavage, all the animals were housed separately and both the water with streptomycin and the food were returned to them. Mice fecal samples were collected for 24 days diluted in PBS, a sample was stored in 15% glycerol at 80ºC and the remaining was plated in Luria Broth agar (LB plates). Plates were incubated overnight at 37ºC and then with the help of a fluorescent stereoscope (SteREO Lumar, Carl Ziess) the fluorescent colonies were counted helping to assess the frequencies of CFP- and YFP-labelled bacteria.

Fifteen mice were gavaged the same way for the first and second colonization, using only different strain as ancestral (see bacterial strains).

### *In vitro* competitions

The evolved clones used for the in vitro competitions were isolated from the last day (day 24 approximately 432 gen) of the two colonizations, one from each population. These clones were subsequently whole genome sequenced and therefore their genotype was determined. To test for fitness advantage of these evolved clones, we competed them with the ancestral of the first step of evolution (5) labelled with a different marker. Three independent competitions were done for each clone. The competition environment was M9 minimal medium (M9-minimal salts, 5x - sigma *) supplemented with either sorbitol (dulcitol, sigma D0256), ribose (D-(-)-Ribose, sigma R7500), mannose (D-(+)-Mannose,99%, Alfa Aesar A10842), gluconate (Gluconic acid

potassium salt, Merck KGaA 8.20601.0500) or glucuronate (D-Glucuronic acid sodium salt monohydrate, sigma G8645) at low concentration (0.02%). Additionally two mixtures with different total amounts of the previous carbon sources (composed of either 0.01% or 0.1% from each of the five carbon sources) were also tested. All competitions were conducted in 96-well plates incubated at 37ºC with aeration (Thermoshaker PHMP-4, Grant). The cultures used for the competitions were previously grown for two overnight growths (first-24h, second-16h) in MM supplemented with glycerol (0.02%). Competitions were performed by inoculating between $10^5$-$10^6$ cells of both competitor and reference strain (in a ~50-50% ratio) in culture medium and allowed to grow for 24 hours. The initial and final ratios of both strains were determined by flow cytometry, using a BD LSRFortessa (BD Biosciences) cytometer. Fitness of the evolved clones was estimated as s (selection coefficient):

$$s = \ln (Nfev / Nfref) / (Niev / Niref) * Gen-1 \qquad Gen = \log2 (Nfref / Niref)$$

Where *Ni* is the initial number of bacteria in the competition, and the *Nf* the final number of bacteria present in the end of the competition. Where *ev* represents the evolved clone and the *ref* the ancestral strain, and the *gen* represents the generation number.

**Statistical analysis**: the data resented in this section represent the means of three experiments ± two standard errors (2se). A Paired with one-tailed distribution *t test* was used to determine the significance of the differences in fitness increase between the two different mixtures(0.05% and 0.5% of carbon source), with a significance defined as P value of <0.05. And a Two-sample unequal variance (heteroscedastic) with one-tailed distribution *t test* was used to determine the significance of the differences in fitness increase between the with the *srlR* mutations and the others when growing with sorbitol as the only carbon source, with a significance defined as P value of <0.05.

*****(33.9g/L $Na_2HPO_47H_2O$+15g/L $KH_2PO_4$ +5g/L $NH_4Cl$ +2.5g/L NaCl))

***In vivo* competitive assays of YFP or CFP sub-populations isolated from the last time point of the first and second colonizations**

Sub-populations of either YFP or CFP clones isolated from the last point of the first and second colonizations, were competed against the respective ancestor (that is either the ancestor of the first or second colonization) labelled with the opposite fluorescent marker. These sub-populations were composed of mixtures of approximately 30 colonies with the same fluorescent marker (population of clones) collected from mouse fecal platings, grown in 10 ml of Luria Broth medium (LB) supplemented with ampicillin (100 µg/ml) or chloramphenicol(100 µg/ml) and Streptomycin(100 µg/ml) and stored in 15% glycerol at -80ºC.

The isolated populations were competed against the ancestral strain labeled with the opposite fluorescent marker, at a ratio of 1 to 1. For the second colonization these numbers were confirmed by flow cytometry, using a BD LSRFortessa (BD Biosciences) cytometer.

After gavage, the animals were housed separately and both food and water containing streptomycin were returned to them. Mice fecal pellets were collected for 3 days, diluted in PBS and frozen in 15% glycerol at -80ºC. Then the final numbers were assessed by fecal plating for the first colonization following the same proceedure described for the evolution experiments and for the second colonization the final numbers were again determined by flow cytometry, using a BD LSRFortessa (BD Biosciences) cytometer.

The selective coefficient (fitness gain) of these clones in vivo (presented in figure 2B) was estimated as:

$$s_b = \ln((Rf_{ev/anc})/(Ri_{ev/anc}))/t$$

Where $s_b$ is the selective coefficient of the evolved clone, $Rf_{ev/anc}$ and $Ri_{ev/anc}$ are the ratios of evolved to ancestral bacteria in the end (f) or in the beginning (i) of the competition an t is the number of generation per day. We assume t=18, in accordance with the 80 minute generation time estimated in previous studies on *E.coli* colonization of streptomycin-treated mouse (86,87)

**Whole genome re-sequencing and mutation prediction**

*Clone analysis*: After 24 days of gut colonization one clone from populations 1.1 to 1.14 from the first colonization, the two ancestors (MG1655-YFP and MG1655-CFP) and one clone from populations 2.1 to 2.15 from the second colonization were isolated and used to seed 10 mL of LB (Line 1.15 was not analyzed since the mouse from this line died at that time point). These cultures were then grown at 37ºC with agitation. Subsequently DNA was isolated following a previously described protocol (Wilson K. 2001). The DNA library construction and sequencing was carried out by BGI. Each sample was pair-end sequenced on an Illumina HiSeq 2000. Standard procedures produced data sets of Illumina paired-end 90 bp read pairs with insert size (including read length) of ,470 bp. Genome sequencing data have been deposited in the NCBI Read Archive, http:// www.ncbi.nlm.nih.gov/sra (accession no. SRP017347). Mutations were identified using the BRESEQ pipeline (88). To detect potential duplication events we used ssaha2 (89) with the paired end information. This is a stringent analysis that maps reads only to their unique match (with less than 3 mismatches) on the reference genome. Sequence coverage along the genome was assessed with a 250 bp window and corrected for GC% composition by normalizing by the mean coverage of regions with the same GC%. We then looked for regions with high differences (>1.4) in coverage. Large deletions were identified based on the absence of coverage. For additional verification of mutations predicted by BRESEQ, we also used the software IGV (version 2.1) (90).

*Ancestral genome*: The sequence reads from MG1655 were mapped to the reference strain (91).The two ancestors carried the mutations listed in Table S1. The mutations underlined were present in the YFP ancestor but not in the CFP. The sequences of the fourteen sequenced clones from the first colonization and the fifteen for the second were then interrogated against this ancestral genome and the mutations identified are listed in Table S2.

*Population analysis*: DNA isolation was obtained in the same way as described above for the clone analysis except that instead of one clone per population a mixture of >1000 clones per population was used. From the second colonization three populations were sequenced: 2.7, 2.10 and 2.14. Populations were sequenced for three time points during the adaptive period (generation 198 (day11), generation 306 (day17) and generation 432 (day24)). From The DNA library construction and sequencing was carried out by the IGC genomics facility. Each sample was pair-end sequenced on an Illumina MiSeq Benchtop Sequencer. Standard procedures produced data sets of Illumina paired-end 250 bp read pairs. The mean coverage per sample was as ~100x for population 2.7, ~100x for population 2.10 and, ~100x for population 2.14. Mutations were identified using the BRESEQ pipeline with the polymorphism option on. The default settings were used except for: a) require a minimum coverage of 3 reads on each strand per polymorphism; b) eliminate polymorphism predictions occurring in homopolymers of length greater than 3; c) eliminate polymorphism predictions with significant ($P$=0.05) strand or base quality score bias.

**Identification of adaptive mutations and estimate of haplotype frequencies in selected populations of the first and second colonizations**

In order to estimate the haplotype frequencies depicted in Fig. 4, depending on the colonization, one or two strategies were employed. For the first colonization target PCR of parallel mutations was performed and for the second colonization, in addition to target PCR, WGS of populations was also obtained. For the target PCR between 20 to 40 clones per time point were isolated. All time points of the first colonization and the last time point of populations 2.7 and 2.14 of the second colonization were analyzed by target PCR. Samples isolated from the populations of the second colonization were WGS in all time points. The isolation procedure consisted in diluting the frozen fecal samples in PBS and plating the appropriate dilution in LB agar plates supplemented with streptomycin (100 µg/ml), and incubating overnight at 37˚C. The frequencies of CFP or YFP bacteria were measured by counting the CFUs in a stereoscope (SteREO Lumar, Carl Zeiss).

Mutations were screened by PCR (see primers in table S6) followed by electrophoresis in 1% agarose gel, at 50V for 1h30min. IS insertions can be scored simply by an increase (around 700-1500bp) in the PCR fragment. *srlR* gene was sequenced using ABI 3130XL and ABI 377 Automatic Sequencer.

The PCR reaction conditions were as follows: one cycle at 95ºC for 3 min followed by 34 cycles of 95 ºC for 30 seg, 60 ºC for 30 seg, 72 ºC for 2 min, followed by 72ºC for 5min. PCR reaction: 1U Taq polymerase,5 µl buffer Taq (10x), 1 µl dNTPs (10mM), 1 µl each primer(Forward and reverse 10 µM), 1 µl of DNA (colony diluted in 10 µl ddH2O) and enough ddH2O to a final volume of 50 µl.

**Test for increased mutation rate**

To test for the possible mutator phenotype of the clone 25YFP ([*insX-insA], yjjP/yjjQ*, *radA* – isolated from population 2.10) (table S2) we determined the frequency of rifampicin-resistant mutants. This was done by growing pre-cultures (10 replica) of the clone 25YFP in 10 ml of LB overnight at 37ºC with aeration. Approximately 1000 cells of the pre-cultures were used to inoculate 1 ml of LB and incubated overnight. Aliquots of each tube were plated in LB agar and LB agar supplemented with rifampicin (100 µg/ml) and incubated overnight at 37ºC. The frequency of spontaneous resistance to rifampicin was calculated as the ratio between the number of rifampicin resistant mutants and the total number of individuals in each plate. The same procedure was applied to the ancestral clone so both rates of the spontaneous resistance to rifampicin could be compared.

**Gene expression**

**Clones**

The clones chosen to measure the effect of the mutations in the genes expression were (for more information see table S1): 18YFP (*focA srlR*), 22YFP (*dcuB*), 25YFP (*yjjP/yjjQ radA insX-insA*), 29CFP (*arcA*) and the ancestral strain of the first colonization. For both aerobic and anaerobic conditions we had three biologic replicates from each clone.

**Growth Conditions**

- **Aerobic Conditions**

After one overnight growth (24h) in MM with glycerol (0.02%) the cultures were diluted (100-fold) and then were grown at 37ºC with aeration, in 10ml of M9 minimal medium (MM) supplemented with a mixture of the five carbon sources (sorbitol, ribose, mannose, gluconate and glucuronate 0.01% from each). At the exponential phase ($OD_{600}$~0.5) we centrifuged 5ml of the bacterial culture at 4ºC during 5minutes at the maximum speed and we discarded the supernatant to prevent its growth. After we added lysozyme solution to disrupt the cells wall (5mg lysozyme/ml DEPC treated water, sigma protocol) and we incubated the samples at 37ºC for 30minutes this process will allow the RNA extraction.

- **Anaerobic conditions**

After one overnight growth (24h) in MM with glycerol (0.02%) the cultures were diluted (100-fold) and then were grown at 37ºC in the anaerobic chamber with the atmosphere of 5% H2, 15% CO2, 80% N2 (Plas Labs, Lansing, MI, USA), in 10ml of M9 minimal medium (MM) supplemented with a mixture of the five carbon sources (sorbitol, ribose, mannose, gluconate and glucuronate 0.01% from each). At the exponential phase ($OD_{600}$~0.2) we positioned the culture tubes in dry ice to prevent its growth. After we centrifuged 5ml of the bacterial culture at 4ºC during 5minutes at the maximum speed and we discarded the supernatant. Next we added lysozyme solution to disrupt the cells wall (5mg lysozyme/ml DEPC treated water, sigma protocol) and we incubated the samples at 37ºC for 30minutes this process will allow the RNA extraction.

**RNA extraction, DNAse treatment, RT-PCR and qPCR**

The RNA extraction was performed with the Quiagen RNeasy Mini Kit. RNA concentration and quality were evaluated with Nanodrop 2000. DNase treatment was performed with the RQ1 DNase (Promega), 0.5ul of DNase and 1ul buffer were added to 1ug of RNA and incubated for 30 minutes at 37ºC. After this, 1ul stop solution was added and then incubated for 15 minutes at 65ºC to inactivate the DNase. The resulting RNA was used for the reverse transcription which consisted in mixing with 1ug of RNA, with 0.5ul random primers and DEPC-water (final volume of 15ul) and then incubate at 70ºC for 5min. Afterwards the M-MLV Reverse Transcriptase Protocol (Promega) were performed, to the first mix were added 5ul of RTbuffer, 0.5ul RT enzyme and 2ul dNTPmix, and then incubated 10min at 25ºC, 50min at 50ºC and 10 min at 70ºC.

We used a relative quantification method of analysis with normalization against a reference gene. qPCR was executed in BioRad CFX 384 with itaqSupermix. cDNA was diluted 100-fold before used in the qPCR. The qPCR reaction conditions were as follows: one cycle of 2 min at 50ºC and then 39cycles of 10 min at 95ºC, 30 sec at 95ºC, 1min at 57ºC and finally 30 s at 72ºC, primers used are listed in table S7. Melt curve analysis was performed to verify product homogeneity. All reactions included three replicates for each sample. Data were normalized by the Pfaffl method (92) using the *hfq* housekeeping gene of *E.coli* as a reference.

**Statistical analysis**: the data presented in this section represent the means of three experiments ± standard errors of the mean (SEM). The unpaired (two-sample equal variance with two-tailed distribution) *t test* was used to determine the significance of the differences in the gene expression between mutant and ancestral strain, with a significance defined as P value of $<0.05$.

## References


1. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the human intestinal microbial flora. Science. 2005 Jun 10;308(5728):1635–8.
2. Grönlund MM, Arvilommi H, Kero P, Lehtonen OP, Isolauri E. Importance of intestinal colonisation in the maturation of humoral immunity in early infancy: a prospective follow up study of healthy infants aged 0-6 months. Arch Dis Child Fetal Neonatal Ed. 2000 Nov;83(3):F186–192.
3. Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. The Placenta Harbors a Unique Microbiome. Sci Transl Med. 2014 May 21;6(237):237ra65–237ra65.
4. Tenaillon O, Skurnik D, Picard B, Denamur E. The population genetics of commensal *Escherichia coli*. Nat Rev Microbiol. 2010 Mar;8(3):207–17.
5. Barroso-Batista J, Sousa A, Lourenço M, Bergman M-L, Sobral D, Demengeot J, et al. The first steps of adaptation of *Escherichia coli* to the gut are dominated by soft sweeps. PLoS Genet. 2014 Mar;10(3):e1004182.
6. Ley RE, Peterson DA, Gordon JI. Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine. Cell. 2006 Feb;124(4):837–48.
7. Blumberg R, Powrie F. Microbiota, Disease, and Back to Health: A Metastable Journey. Sci Transl Med. 2012 Jun 6;4(137):137rv7–137rv7.
8. O'Hara AM, Shanahan F. The gut flora as a forgotten organ. EMBO Rep. 2006 Jul;7(7):688–93.
9. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010 Mar 4;464(7285):59–65.
10. Weinstock GM. Genomic approaches to studying the human microbiota. Nature. 2012 Sep 13;489(7415):250–6.
11. Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. Nat Rev Microbiol. 2011 Apr;9(4):279–90.
12. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A. 1998 Jun 9;95(12):6578–83.
13. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature. 2011 May 12;473(7346):174–80.
14. Sekirov I, Russell SL, Antunes LCM, Finlay BB. Gut Microbiota in Health and Disease. Physiol Rev. 2010 Jul 27;90(3):859–904.
15. Escherich T. The intestinal bacteria of the neonate and breast-fed infant. 1885. Rev Infect Dis. 1989 Apr;11(2):352–6.
16. Savageau MA. *Escherichia coli* Habitats, Cell Types, and Molecular Mechanisms of Gene Control. The American Naturalist. 1983 Dec;
17. Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, et al. Factors influencing the composition of the intestinal microbiota in early infancy. Pediatrics. 2006 Aug;118(2):511–21.
18. Gordon DM. The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. Microbiology. 2003 Dec 1;149(12):3575–86.
19. Magne F, Suau A, Pochart P, Desjeux J-F. Fecal microbial community in preterm infants. J Pediatr Gastroenterol Nutr. 2005 Oct;41(4):386–92.
20. Caugant DA, Levin BR, Selander RK. Genetic diversity and temporal variation in the E. coli population of a human host. Genetics. 1981 Jul;98(3):467–90.
21. Sears HJ, Brownlee I, Uchiyama JK. Persistence of individual strains of *Escherichia coli* in the intestinal tract of man. J Bacteriol. 1950 Feb;59(2):293–301.
22. Hendrickson H. Order and Disorder during *Escherichia coli* Divergence. Casadesús J, editor. PLoS Genet. 2009 Jan 23;5(1):e1000335.
23. Herzer PJ, Inouye S, Inouye M, Whittam TS. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. J Bacteriol. 1990 Nov;172(11):6175–81.
24. Kaper JB, Nataro JP, Mobley HLT. Pathogenic *Escherichia coli*. Nat Rev Microbiol. 2004 Feb;2(2):123–40.
25. Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev. 1998 Jan;11(1):142–201.
26. Donnenberg MS, Whittam TS. Pathogenesis and evolution of virulence in enteropathogenic and enterohemorrhagic *Escherichia coli*. J Clin Invest. 2001 Mar;107(5):539–48.

27. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, et al. The Long-Term Stability of the Human Gut Microbiota. Science. 2013 Jul 5;341(6141):1237439–1237439.
28. David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, et al. Host lifestyle affects human microbiota on daily timescales. Genome Biol. 2014 Jul 25;15(7):R89.
29. Turnbaugh PJ, Bäckhed F, Fulton L, Gordon JI. Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome. Cell Host Microbe. 2008 Apr;3(4):213–23.
30. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, et al. Evolution of Mammals and Their Gut Microbes. Science. 2008 Jun 20;320(5883):1647–51.
31. Russo TA, Johnson JR. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. Microbes Infect Inst Pasteur. 2003 Apr;5(5):449–56.
32. Isolauri E, Salminen S, Ouwehand AC. Probiotics. Best Pract Res Clin Gastroenterol. 2004 Apr;18(2):299–313.
33. Hooper LV, Littman DR, Macpherson AJ. Interactions Between the Microbiota and the Immune System. Science. 2012 Jun 8;336(6086):1268–73.
34. Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, et al. Host-Gut Microbiota Metabolic Interactions. Science. 2012 Jun 8;336(6086):1262–7.
35. Stecher B, Berry D, Loy A. Colonization resistance and microbial ecophysiology: using gnotobiotic mouse models and single-cell technology to explore the intestinal jungle. FEMS Microbiol Rev. 2013 Sep;37(5):793–829.
36. Gibson GR, Probert HM, Loo JV, Rastall RA, Roberfroid MB. Dietary modulation of the human colonic microbiota: updating the concept of prebiotics. Nutr Res Rev. 2004 Dec;17(2):259–75.
37. Vieira AT, Teixeira MM, Martins FS. The Role of Probiotics and Prebiotics in Inducing Gut Immunity. Front Immunol [Internet]. 2013 [cited 2014 Sep 18];4. Available from: http://journal.frontiersin.org/Journal/10.3389/fimmu.2013.00445/full
38. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006 Dec 21;444(7122):1027–131.
39. Khoruts A, Weingarden AR. Emergence of fecal microbiota transplantation as an approach to repair disrupted microbial gut ecology. Immunol Lett [Internet]. 2014 Aug [cited 2014 Sep 18]; Available from: http://linkinghub.elsevier.com/retrieve/pii/S0165247814001564
40. Song Y, Garg S, Girotra M, Maddox C, von Rosenvinge EC, Dutta A, et al. Microbiota Dynamics in Patients Treated with Fecal Microbiota Transplantation for Recurrent Clostridium difficile Infection. Berg G, editor. PLoS ONE. 2013 Nov 26;8(11):e81330.
41. Van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, de Vos WM, et al. Duodenal Infusion of Donor Feces for Recurrent *Clostridium difficile*. N Engl J Med. 2013 Jan 31;368(5):407–15.
42. Gill SR. Metagenomic Analysis of the Human Distal Gut Microbiome. Science. 2006 Jun 2;312(5778):1355–9.
43. Freter R, Brickner H, Botney M, Cleven D, Aranki A. Mechanisms that control bacterial populations in continuous-flow culture models of mouse large intestinal flora. Infect Immun. 1983 Feb;39(2):676–85.
44. Chang D-E. Carbon nutrition of *Escherichia coli* in the mouse intestine. Proc Natl Acad Sci. 2004 May 5;101(19):7427–32.
45. Peekhaus N, Conway T. What's for dinner?: Entner-Doudoroff metabolism in *Escherichia coli.* J Bacteriol. 1998 Jul;180(14):3495–502.
46. Hoskins LC, Agustines M, McKee WB, Boulding ET, Kriaris M, Niedermeyer G. Mucin degradation in human colon ecosystems. Isolation and properties of fecal strains that degrade ABH blood group antigens and oligosaccharides from mucin glycoproteins. J Clin Invest. 1985 Mar;75(3):944–53.
47. Cooper TF, Lenski RE. Experimental evolution with E. coli in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. BMC Evol Biol. 2010;10(1):11.
48. Habets MGJL, Rozen DE, Hoekstra RF, de Visser JAGM. The effect of population structure on the adaptive radiation of microbial populations evolving in spatially structured environments. Ecol Lett. 2006 Sep;9(9):1041–8.

49. Quan S, Ray JCJ, Kwota Z, Duong T, Balázsi G, Cooper TF, et al. Adaptive evolution of the lactose utilization network in experimentally evolved populations of *Escherichia coli*. PLoS Genet. 2012 Jan;8(1):e1002444.
50. Kinnersley M, Wenger J, Kroll E, Adams J, Sherlock G, Rosenzweig F. Ex Uno Plures: Clonal Reinforcement Drives Evolution of a Simple Microbial Community. Zhang J, editor. PLoS Genet. 2014 Jun 26;10(6):e1004430.
51. Zhong S, Miller SP, Dykhuizen DE, Dean AM. Transcription, Translation, and the Evolution of Specialists and Generalists. Mol Biol Evol. 2009 Aug 25;26(12):2661–78.
52. He G, Shankar RA, Chzhan M, Samouilov A, Kuppusamy P, Zweier JL. Noninvasive measurement of anatomic structure and intraluminal oxygenation in the gastrointestinal tract of living mice with spatial and spectral EPR imaging. Proc Natl Acad Sci U S A. 1999 Apr 13;96(8):4586–91.
53. Madigan MT, Martinko JM, Dunlap PV, Clark DP. Brock Biology of Microorganisms. Twelfth edition. Pearson Benjamin Cummings; 2009.
54. Jones SA, Chowdhury FZ, Fabich AJ, Anderson A, Schreiner DM, House AL, et al. Respiration of *Escherichia coli* in the Mouse Intestine. Infect Immun. 2007 Aug 13;75(10):4891–9.
55. De Paepe M, Gaboriau-Routhiau V, Rainteau D, Rakotobe S, Taddei F, Cerf-Bensussan N. Trade-Off between Bile Resistance and Nutritional Competence Drives *Escherichia coli* Diversification in the Mouse Gut. Casadesús J, editor. PLoS Genet. 2011 Jun 16;7(6):e1002107.
56. Lassig M. Chance and risk in adaptive evolution. Proc Natl Acad Sci. 2012 Mar 12;109(13):4719–20.
57. Gordo I, Perfeito L, Sousa A. Fitness Effects of Mutations in Bacteria. J Mol Microbiol Biotechnol. 2011;21(1-2):20–35.
58. Mahillon J, Chandler M. Insertion sequences. Microbiol Mol Biol Rev MMBR. 1998 Sep;62(3):725–74.
59. Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, et al. Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. Nat Genet. 2003 Sep;35(1):32–40.
60. Casacuberta E, González J. The impact of transposable elements in environmental adaptation. Mol Ecol. 2013 Mar;22(6):1503–17.
61. Barrick JE, Lenski RE. Genome dynamics during experimental evolution. Nat Rev Genet. 2013 Oct 29;14(12):827–39.
62. Perfeito L, Fernandes L, Mota C, Gordo I. Adaptive mutations in bacteria: high rate and small effects. Science. 2007 Aug 10;317(5839):813–5.
63. Messer PW, Petrov DA. Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol Evol. 2013 Nov;28(11):659–69.
64. Herron MD, Doebeli M. Parallel Evolutionary Dynamics of Adaptive Diversification in *Escherichia coli*. Nosil P, editor. PLoS Biol. 2013 Feb 19;11(2):e1001490.
65. Maharjan RP, Ferenci T, Reeves PR, Li Y, Liu B, Wang L. The multiplicity of divergence mechanisms in a single evolving population. Genome Biol. 2012;13(6):R41.
66. Rainey null, Buckling null, Kassen null, Travisano null. The emergence and maintenance of diversity: insights from experimental bacterial populations. Trends Ecol Evol. 2000 Jun;15(6):243–7.
67. Hentges DJ, Pongpech P, Que JU. Hypothesis: How Streptomycin Treatment Compromises Colonisation Resistance Against Enteric Pathogens in Mice. Microb Ecol Health Dis [Internet]. 1990 Jun 1 [cited 2014 Sep 22];3(3). Available from: http://www.microbecolhealthdis.net/index.php/mehd/article/view/7522
68. Leatham-Jensen MP, Frimodt-Moller J, Adediran J, Mokszycki ME, Banner ME, Caughron JE, et al. The Streptomycin-Treated Mouse Intestine Selects *Escherichia coli* envZ Missense Mutants That Interact with Dense and Diverse Intestinal Microbiota. Infect Immun. 2012 May 1;80(5):1716–27.
69. Hegreness M, Shoresh N, Hartl D, Kishony R. An equivalence principle for the incorporation of favorable mutations in asexual populations. Science. 2006 Mar 17;311(5767):1615–7.
70. Nobelmann B, Lengeler JW. Molecular analysis of the gat genes from *Escherichia coli* and of their roles in galactitol transport and metabolism. J Bacteriol. 1996 Dec;178(23):6790–5.
71. Jones SA, Gibson T, Maltby RC, Chowdhury FZ, Stewart V, Cohen PS, et al. Anaerobic Respiration of *Escherichia coli* in the Mouse Intestine. Infect Immun. 2011 Aug 8;79(10):4218–26.

72. Stoebel DM, Dean AM, Dykhuizen DE. The Cost of Expression of *Escherichia coli* lac Operon Proteins Is in the Process, Not in the Products. Genetics. 2008 Feb 1;178(3):1653–60.
73. Zhong S, Khodursky A, Dykhuizen DE, Dean AM. Evolutionary genomics of ecological specialization. Proc Natl Acad Sci U S A. 2004 Aug 10;101(32):11719–24.
74. Suppmann B, Sawers G. Isolation and characterization of hypophosphite--resistant mutants of *Escherichia coli*: identification of the FocA protein, encoded by the pfl operon, as a putative formate transporter. Mol Microbiol. 1994 Mar;11(5):965–82.
75. Salmon K, Hung S -p., Mekjian K, Baldi P, Hatfield GW, Gunsalus RP. Global Gene Expression Profiling in *Escherichia coli* K12: THE EFFECTS OF OXYGEN AVAILABILITY AND FNR. J Biol Chem. 2003 Aug 8;278(32):29837–55.
76. Marzan LW, Siddiquee KAZ, Shimizu K. Metabolic regulation of an fnr gene knockout *Escherichia coli* under oxygen limitation. Bioeng Bugs. 2011 Nov 1;2(6):331–7.
77. Chou H-H, Chiu H-C, Delaney NF, Segre D, Marx CJ. Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation. Science. 2011 Jun 3;332(6034):1190–2.
78. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N Engl J Med. 1994 Jun 16;330(24):1703–9.
79. Stanley J, Baquar N, Threlfall EJ. Genotypes and phylogenetic relationships of *Salmonella typhimurium* are defined by molecular fingerprinting of IS200 and 16S rrn loci. J Gen Microbiol. 1993 Jun;139 Pt 6:1133–40.
80. Bik EM, Gouw RD, Mooi FR. DNA fingerprinting of Vibrio cholerae strains with a novel insertion sequence element: a tool to identify epidemic strains. J Clin Microbiol. 1996 Jun;34(6):1453–61.
81. Elena SF, Lenski RE. Microbial genetics: Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet. 2003 Jun;4(6):457–69.
82. Autieri SM, Lins JJ, Leatham MP, Laux DC, Conway T, Cohen PS. L-Fucose Stimulates Utilization of D-Ribose by *Escherichia coli* MG1655 fucAO and E. coli Nissle 1917 fucAO Mutants in the Mouse Intestine and in M9 Minimal Medium. Infect Immun. 2007 Nov 1;75(11):5465–75.
83. Cooper VS, Lenski RE. The population genetics of ecological specialization in evolving *Escherichia col*i populations. Nature. 2000 Oct 12;407(6805):736–9.
84. Good BH, Rouzine IM, Balick DJ, Hallatschek O, Desai MM. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. Proc Natl Acad Sci. 2012 Mar 27;109(13):4950–5.
85. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of *Escherichia coli* K-12. Science. 1997;277(5331):1453–62.
86. Poulsen LK, Licht TR, Rang C, Krogfelt KA, Molin S. Physiological state of *Escherichia col*i BJ4 growing in the large intestines of streptomycin-treated mice. J Bacteriol. 1995 Oct;177(20):5840–5.
87. Rang CU, Licht TR, Midtvedt T, Conway PL, Chao L, Krogfelt KA, et al. Estimation of growth rates of *Escherichia col*i BJ4 in streptomycin-treated and previously germfree mice by in situ rRNA hybridization. Clin Diagn Lab Immunol. 1999 May;6(3):434–6.
88. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, et al. Genome evolution and adaptation in a long-term experiment with *Escherichia col*i. Nature. 2009 Oct 29;461(7268):1243–7.
89. Ning Z, Cox AJ, Mullikin JC. SSAHA: a fast search method for large DNA databases. Genome Res. 2001 Oct;11(10):1725–9.
90. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011 Jan;29(1):24–6.
91. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of *Escherichia coli* K-12. Science. 1997 Sep 5;277(5331):1453–62.
92. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001 May 1;29(9):e45.

## Supplements

Table S1 – Mutations identified in the genomes of the ancestral strains of the first colonization (53). Mutations in intergenic regions have the two flanking genes listed (e.g., fdrA/ylbF). Genes within brackets mean that the mutation happened within the gene. SNPs are represented by an arrow between the ancestral and the evolved nucleotide. Whenever a SNP gives rise to a non-synonymous mutation the amino acid replacement is also indicated. The symbol Δ means a deletion event and a + symbol represents an insertion of the nucleotide that follows the symbol. For intergenic mutations, the numbers in the annotation row represent nucleotides relative to each of the neighboring genes, where + indicates the distance downstream of the stop codon of a gene and - indicates the distance upstream of the gene, that is relative to the start codon. The mutations present in the 0YFP but not in the 0CFP, are underlined.

| Clone | Genome Position | Gene | Mutation | Annotation |
|---|---|---|---|---|
| **0YFP** | 547694 | *fdrA/ylbF* | A→G | intergenic (+123/-1156) |
| | 547835 | *fdrA/ylbF* | +G | intergenic (+264/-1015) |
| | 1395405 | *[ynaJ]–[ttcA]* | Δ13,756 bp | multigenic |
| | 1976527 | *insB–insA* | Δ776 bp | |
| | 2369558 | *arnT* | 4 bp x 2 | duplication |
| | 3422257 | *rrlD* | A→C | noncoding |
| | 3422258 | *rrlD* | T→A | noncoding |
| | 3422259 | *rrlD* | C→T | noncoding |
| | 3434719 | *trkA* | G→A | E60E GAG→GAA |
| | 3472447 | *rpsL* | T→C | K43R AAA→AGA |
| | 3844290 | *uhpT* | A→C | F301V TTT→GTT |
| | 3957957 | *ppiC/rep* | C→T | intergenic (-121/-743) |
| | 4095684 | *rhaB/rhaS* | T→C | intergenic (-213/-75) |
| | 360473 | *lacA–lacI* | Δ6264 bp | multigenic |
| | 788169 | *[galK]* | Δ1034 bp | |
| | 4294082 | *RIP321* | Δ338 bp | |

**Table S2.1** -Number and nature of adaptive events across independently evolved clones from the first colonization to the mouse gut (53). In the list of mutations, the initials IS denote the abbreviation of insertion sequence element at the indicated position. The asterisk means that the corresponding SNP originated a STOP codon. For further details see Table S1 legend. The last column shows the number of mutations segregating in each sequenced clone. Clone 12YFP was the ancestral strain to the second colonization. Mutation *yjjY/yjjD* is in the *arcA* regulatory region.

| First Step of Adaptation to the mouse gut | | | | | |
|---|---|---|---|---|---|
| **Clone** | **Genome Position** | **Gene** | **Mutation** | **Annotation** | **Inferred mutations** |
| **1CFP** | 2173589 | *gatZ* | IS Ins | coding (755/1263) | 2 |
| | 4346888 | *dcuB/dcuR* | IS Ins | intergenic (-121/+450) | |
| **2YFP** | 2173759 | *gatZ* | IS Ins | coding (585/1263) | 2 |
| | 4346888 | *dcuB/dcuR* | IS Ins | intergenic (-121/+450) | |
| **3YFP** | 2560011 | *yffN* | G→A | C122Y TGC→TAC | 3 |
| | 2175242 | *gatY/fbaB* | IS Ins | intergenic (-16/+292) | |
| | 4601260 | *yjjP/yjjQ* | IS Ins | intergenic (-379/-244) | |
| **4YFP** | 2173531 | *gatZ* | IS1 Ins | coding (813/1263) | 1 |
| **5YFP** | 2827493 | *srlR* | G→C | G142A GGC→GCC | 2 |
| | 2172869 | *gatA* | IS Ins | coding (203/453) | |
| **6YFP** | 2172262 | *gatC* | del 1 bp | coding (39/1356) | 2 |
| | 4346888 | *dcuB/dcuR* | IS Ins | intergenic (-121/+450) | |
| **7YFP** | 1420379 | *ydaV* | C→A | L125M CTG→ATG | 4 |
| | 1902231 | *[manZ]–[kdgR]* | Δ5,451 bp | | |
| | 2175298 | *gatY/fbaB* | IS Ins | intergenic (-72/+236) | |
| | 953904 | *focA/ycaO* | IS Ins | intergenic (-212/+194) | |
| **8YFP** | 3268729 | *garK* | A→G | F355S TTC→TCC | 2 |
| | 2172869 | *gatA* | IS Ins | coding (203/453) | |
| **9YFP** | 2827073 | *srlR* | A→T | K2I AAA→ATA | 2 |
| | 2172636 | *gatA* | IS Ins | coding (436/453) | |
| **10CFP** | 2174223 | *gatZ* | Δ2 bp | coding (120-121/1263) | 3 |
| | 1388754 | *[ycjY-ynaI]* | Δ5315 bp | large deletion | |
| | 953904 | *focA/ycaO* | IS5 Ins | intergenic (-212/+194) | |
| **11CFP** | 2827095 | *srlR* | Δ1 bp | coding (27/774) | 2 |
| | 2172869 | *gatA* | IS Ins | coding (203/453) | |
| **12YFP** | 2172079 | *gatC* | +C | coding (222/1356) | 2 |
| | 4500113 | *[insG-yaaI]* | 2x 151716 bp | large duplication | |
| **13CFP** | 4637714 | *arcA* | G→T | R206S CGC→AGC | 2 |
| | 2171153 | *gatC* | IS Ins | coding (1148/1356) | |
| **14CFP** | 943941 | *dmsC* | G→A | W229* TGG→TAG | 3 |
| | 2175263 | *gatY/fbaB* | IS Ins | intergenic (-37/+271) | |
| | 2827117 | *srlR* | C→T | Q17* CAG→TAG | |

**Table S2.2** -Number and nature of adaptive events across independently evolved clones from the second colonization to the mouse gut.

| Second Step of Adaptation to the mouse gut | | | | | |
|---|---|---|---|---|---|
| **Clone** | **Genome Position** | **Gene** | **Mutation** | **Annotation** | **Inferred mutations** |
| **16YFP** | =17531<br>4500113 | nhaA<br>noncoding (1426/1426 nt) | Duplication between 4500113 and 17531 | coding (43/1167 nt)<br>IS4 | 2 |
| **17YFP** | 4638771 | *yjjY/yjtD* | IS5 (–) +4 bp | intergenic (+206/-191) | 1 |
| **18YFP** | 953901<br>2827229 | *focA/ycaO*<br>*srlR* | IS5 (+) +4 bp<br>G→A | intergenic (-212/+191)<br>G54D (GGC→GAC) | 2 |
| **19CFP** | 317172<br>2556721 | *ykgB*<br>*intZ–[eutA]* | IS2 (–) +5 bp<br>Δ6,790 bp | coding (368-372/594 nt) | 2 |
| **20CFP** | 4347121 | *dcuB/dcuR* | A→T | intergenic (-354/+217) | 1 |
| **21YFP** | 4601066<br>4638684 | *yjjP/yjjQ*<br>*yjjY/yjtD* | IS2 (–) +5 bp<br>+TTAT | intergenic (-185/-430)<br>intergenic (+119/-281) | 2 |
| **22YFP** | 4347105 | *dcuB/dcuR* | IS2 (–) +5 bp | intergenic (-338/+229) | 1 |
| **23CFP** | 3174801<br>4638719 | *cpdA*<br>*yjjY/yjtD* | G→A<br>C→A | Q19* (CAA→TAA)<br>intergenic (+154/-246) | 2 |
| **24YFP** | 4625061 | *radA* | IS1 (–) +9 bp | coding (1127-1135/1383 nt) | 1 |
| **25YFP** | 279155<br>4600992<br>4625061 | *insX–insA*<br>*yjjP/yjjQ*<br>*radA* | Δ11,471 bp<br>IS1 (–) +9 bp<br>IS1 (–) +9 bp | between IS1<br>intergenic (-111/-500)<br>coding (1127-1135/1383 nt) | 3 |
| **26YFP** | 1859452<br>4346885 | *ydjL/yeaC*<br>*dcuB/dcuR* | G→T<br>IS5 (–) +4 bp | intergenic (-96/+274)<br>intergenic (-118/+450) | 2 |
| **27CFP** | 1301748<br>2827045 | *oppB*<br>*gutM/srlR* | IS5 (–) +4 bp<br>A→T | coding (826-829/921 nt)<br>intergenic (+43/-24) | 2 |
| **28YFP** | 1301338<br>2827745<br>3805903 | *oppB*<br>*srlR*<br>*rfaQ* | IS186 (+) +6 bp :: Δ1<br>+T<br>IS1 (–) +9 bp | coding (416-421/921 nt)<br>coding (677/774 nt)<br>coding (211-219/1035 nt) | 3 |
| **29CFP** | 4638628 | *yjjY/yjtD* | IS2 (–) +5 bp | intergenic (+63/-333) | 1 |
| **30YFP** | 9531240<br>4031240 | *focA/ycaO*<br>*trkH* | IS1(-)+9bp<br>G→A | intergenic (-174/+224)<br>G275R GGG→AGG | 2 |

**Table S3** - Frequencies of newly generated mutations along 432 generations of evolution of population 2.7 inside the mouse gut. Clone 12 YFP (see table S2) was the ancestral of the second colonization. Yellow and blue shading identify haplotypes belonging to the two sub-populations of bacteria labeled either with *cfp* or *yfp* alleles, the ones with no shading are the mutations whose background and haplotype was not possible to estimate. The ancestral haplotype is, by definition, devoid of mutations. The genome position for a given mutation is only indicated for mutations that are not IS insertions. Mutations in intergenic regions have the two flanking genes listed (e.g., *dcuB/dcuR*). The symbol Δ means a deletion event and a + symbol represents an insertion of the nucleotide that follows the symbol. IS insertions at given position are indicated as IS Ins. See Figure 4 for a graphical representation of the data in this table.

| Genome position | gene | mutation | Haplotype/mutation frequencies | | | |
|---|---|---|---|---|---|---|
| | | | d0 | 198(d11) | 306(d17) | 432(d24) |
| | | | 0.5 | 0.428 | 0.447 | 0.143 |
| | *arcA* | IS ins | 0 | 0.245 | 0.311 | 0.48 |
| | *dcuB* | IS ins | 0 | 0 | 0.07 | 0.23 |
| | | | 0.5 | 0.327 | 0.172 | 0.147 |
| | *dcuB* | IS ins | 0 | 0 | 0 | 0.05 |
| | *arcA* | IS ins | 0 | 0 | 0.07 | 0 |
| 2827065 | *srlR(+C)* | +C | 0 | 0.013 | 0 | 0 |
| 2827066 | *srlR(C-.)* | C - . | 0 | 0 | 0 | 0.018 |
| | *dcuB* | IS ins | 0 | 0 | 0 | 0.08 |
| 4346647 | *dcuB* | G-. | 0 | 0 | 0.04 | 0 |
| 2302936 | *eco/mqo* | A-G | 0 | 0.067 | 0 | 0 |
| 2858252 | *pphB* | +G | 0 | 0.039 | 0 | 0 |
| 251816 | *dinB* | Δ48bp | 0 | 0.085 | 0 | 0 |
| 1131134 | *flgD* | | 0 | 0.036 | 0 | 0 |
| 34112 | *carB/caiF* | Δ1 bp | 0 | 0 | 0.05 | 0 |

**Table S4** - Frequencies of newly generated haplotypes along 432 generations of evolution of population 2.14 inside the mouse gut. See Table S3 for further details.

| Genome Position | gene | mutation | Haplotype/mutations frequencies | | | |
|---|---|---|---|---|---|---|
| | | | d0 | 198(d11) | 306(d17) | 432(d24) |
| | | | 0.5 | 0.607 | 0.528 | 0.273 |
| | *arcA* | IS ins | 0 | 0 | 0 | 0.05 |
| | *focA/ycaO* | IS ins | 0 | 0 | 0 | 0.15 |
| | | | 0.5 | 0.393 | 0.472 | 0.477 |
| | *arcA* | IS ins | 0 | 0 | 0 | 0.05 |
| | *yjjP/yjjQ* | IS ins | 0 | 0 | 0.35 | 0 |
| 2172089 | *gatC* | Δ1 bp | 0 | 0 | 0.129 | 0 |
| 1740247 | *cfa* | Δ1 bp | 0 | 0 | 0 | 0.057 |
| 3718679 | *hokA/insJ* | C-T | 0 | 0 | 0 | 0.068 |

**Table S5** - Frequencies of newly generated haplotypes along 432 generations of evolution of population 2.10 inside the mouse gut. See Table S3 for further details.

| Genome Position | gene | mutation | Haplotype/mutations frequencies | | | |
|---|---|---|---|---|---|---|
| | | | d0 | 198(d11) | 306(d17) | 432(d24) |
| | | | 0.5 | 0 | 0 | 0 |
| | *radA* | IS ins | 0 | 0.8689 | 0.707 | 0 |
| | *radA*<br>*yjjP/yjjQ* | IS ins<br>IS ins | 0 | 0 | 0.065-<br>0.095 | 0.66 |
| | *radA*<br>*yjjP/yjjQ* | IS ins<br>IS ins | 0 | 0 | 0 | 0.29-0.35 |
| | *radA*<br>*focA/ycaO* | IS ins<br>IS ins | 0 | 0 | 0.143 | 0 |
| | *radA*<br>*focA/ycaO* | IS ins<br>IS ins | 0 | 0 | 0.1 | 0 |
| 2827162 | *radA*<br>*srlR* | IS ins<br>A-G | 0 | 0 | 0 | 0.057 |
| 2827271 | *radA*<br>*srlR* | IS ins<br>.-T | 0 | 0 | 0 | 0.016 |
| | *radA*<br>*yciG/trpA* | IS ins<br>IS ins | 0 | 0.1311 | 0 | 0 |
| 1636737 | *radA*<br>*cspI/rzpQ* | IS ins<br>Δ1 bp | 0 | 0 | 0.05 | 0 |
| 301 | *radA*<br>*thrL/thrA* | IS ins<br>.-T | 0 | 0 | 0 | 0.023 |
| 379237 | *radA*<br>*frmR/yaiO* | IS ins<br>Δ1 bp | 0 | 0 | 0 | 0.125 |
| | *radA*<br>*mngB/cydA* | IS ins<br>IS ins | 0 | 0 | 0 | 0.084 |
| | | | 0.5 | 0 | 0 | 0 |

**Table S6** –Oligonucleotide primers used in PCR.

| | | |
|---|---|---|
| srlR | forward | GCATGCGGGTGATTTACAGC |
| | reverse | TTCCGGTAAACGGCTTGCTT |
| dcuB | forward | GGCTGAAGGTGGAAGACGAA |
| | reverse | ACATTTCGCGTGTTTCCTGC |
| focA | forward | AGCGGATGTTTCGTTGCTTT |
| | reverse | TGCTGCACATCAGTCGTTGT |
| yjjP/yjjQ | forward | GCTGGACTTCAGTCACCACA |
| | reverse | TTGCAGCATCCTGGCAACATA |
| arcA | forward | AAAAGTCCCTGACCTGCCTG |
| | reverse | TGGCGACAGTGAAATCGACA |

**Table S7** – Oligonucleotide primers used in qPCR

| | | |
|---|---|---|
| dcuB | forward | TGCGATACCGTACACCGCCAC |
| | reverse | CATCCTTCGGCGGCAAACCG |
| focA | forward | GGTGCACTTTGTGGTCGGCG |
| | reverse | CACCTGGGGTCAGTTGGCGA |
| yjjP | forward | ACGGAAAGCCCGGCACTAGC |
| | reverse | TCGCCGCCACCACCATTTCC |
| arcA | forward | CGTGCGGTTTCAGCTCACGG |
| | reverse | CGAGTTCCGCGCCATGCTTC |
| yjjY | forward | TTGCGTTCTTGATGCACTTTCCA |
| | reverse | ACACTGTCGGGTCCTGAGGGA |
| srlA | forward | AACGTTGCGCCGGTAACCCT |
| | reverse | AAGAGGCCGCCGCGTAGTAG |
| hfq | forward | AGCACGCGATTTCTACTGTTGTCC |
| | reverse | CACCGGCGTTGTTACTGTGATG |

**Figure S1 –** Emergence and spread of beneficial mutations in the *gat* operon. Dynamics of frequency change of the gat-negative phenotype over time for all populations from the first colonization (1.1 to 1.5). For more information see Barroso-Batista et al 2014 (5)**.**